How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans

Alan Cartmell[a,1], Elisabeth C. Lowe[a,1], Arnaud Baslé[a], Susan J. Firbank[a], Didier A. Ndeh[a], Heath Murray[a], Nicolas Terrapon[b], Vincent Lombard[b], Bernard Henrissat[b,c,d], Jeremy E. Turnbull[e], Mirjam Czjzek[f,g], Harry J. Gilbert[a], and David N. Bolam[a,2]

[a]Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, United Kingdom; [b]Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille University, F-13288 Marseille, France; [c]Institut National de la Recherche Agronomique, USC1408 Architecture et Fonction des Macromolécules Biologiques, F-13288 Marseille, France; [d]Department of Biological Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia; [e]Centre for Glycobiology, Department of Biochemistry, Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, United Kingdom; [f]Sorbonne Universités, Université Pierre-et-Marie-Curie, Université Paris 06, F-29688 Roscoff cedex, Bretagne, France; and [g]CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, F-29688 Roscoff cedex, Bretagne, France

Edited by Carolyn R. Bertozzi, Stanford University, Stanford, CA, and approved May 23, 2017 (received for review March 17, 2017)

The human microbiota, which plays an important role in health and disease, uses complex carbohydrates as a major source of nutrients. Utilization hierarchy indicates that the host glycosaminoglycans heparin (Hep) and heparan sulfate (HS) are high-priority carbohydrates for *Bacteroides thetaiotaomicron*, a prominent member of the human microbiota. The sulfation patterns of these glycosaminoglycans are highly variable, which presents a significant enzymatic challenge to the polysaccharide lyases and sulfatases that mediate degradation. It is possible that the bacterium recruits lyases with highly plastic specificities and expresses a repertoire of enzymes that target substructures of the glycosaminoglycans with variable sulfation or that the glycans are desulfated before cleavage by the lyases. To distinguish between these mechanisms, the components of the *B. thetaiotaomicron* Hep/HS degrading apparatus were analyzed. The data showed that the bacterium expressed a single-surface endo-acting lyase that cleaved HS, reflecting its higher molecular weight compared with Hep. Both Hep and HS oligosaccharides imported into the periplasm were degraded by a repertoire of lyases, with each enzyme displaying specificity for substructures within these glycosaminoglycans that display a different degree of sulfation. Furthermore, the crystal structures of a key surface glycan binding protein, which is able to bind both Hep and HS, and periplasmic sulfatases reveal the major specificity determinants for these proteins. The locus described here is highly conserved within the human gut *Bacteroides*, indicating that the model developed is of generic relevance to this important microbial community.

human gut microbiota | glycosaminoglycan degradation | heparin | heparan sulfate | *Bacteroides thetaiotaomicron*

The human colonic microbiota (CM) is crucial to health (1–3). The composition of the CM depends on its ability to access nutrients, which are primarily dietary and host glycans. Dissecting the mechanisms by which complex carbohydrates are used by the CM is critical to understanding the drivers of the ecology of this microbial community and how this process relates to human health.

The major glycan degraders in the CM are the Bacteroidetes (4–6). These organisms access their target polysaccharides through endo-acting enzymes on the bacterial surface followed by import of the oligosaccharides generated, which are depolymerized in the periplasm. The genes encoding these enzyme systems are physically linked into loci termed polysaccharide utilization loci (PULs) (7). A significant proportion of the complex carbohydrates available to the CM is mammalian in origin (4). Despite this knowledge, our understanding of how the CM accesses host/mammalian glycans is fragmentary. Models for the breakdown of high-mannose and complex *N*-glycans by gut *Bacteroides* have been proposed (8), and the ecological significance of removing terminal sialic acid from the epithelial mucosa is established (9). There is, however, a paucity of information on the mechanism by which glycosaminoglycans (GAGs), such as heparin (Hep) and heparan sulfate (HS), are used by the CM. HS is a major component of the extracellular matrix of mammalian cells and therefore, likely to be available to the gut microbiota via turnover of epithelial cells, whereas Hep is released from mast cells at sites of injury and therefore, may not be as prevalent as HS in the gut (10, 11).

Microbial utilization of Hep and HS poses significant biological challenges. Both glycans differ significantly in their degree of polymerization (DP), which suggests that degradation may occur in different cellular locations, whereas sulfation patterns and the uronic acid (UA) are also variable (Fig. 1) (10). This substantial heterogeneity indicates that either a complex portfolio of enzymes is required to deconstruct these acidic glycans or the enzymes that mediate this process display significant substrate promiscuity. The depolymerization of UA-containing glycans, such as Hep/HS, is mediated by glycoside hydrolases (GHs) and/or polysaccharide lyases (PLs) that are grouped into sequence-based families in the CAZy database (12). Hep/HS are degraded by bacterial PLs belonging to families PL12, -13, and -21. Based on specificity, these PLs can be further broadly grouped into three functional groups,

Significance

The major nutrients available to the human microbiota are complex carbohydrates. Host glycans are important to this microbial community, particularly when dietary carbohydrates are scarce. The host glycans heparin and heparan sulfate are high-priority carbohydrates for *Bacteroides thetaiotaomicron*, a member of the human microbiota. The degradation of these complex carbohydrates is challenging, reflecting their highly variable sulfation patterns. How bacteria have adapted to depolymerize the myriad of substructures of this important class of glycosaminoglycan is unknown. Here, we show how enzyme consortia, displaying complementary functions, target the different features of these host glycans. Structural data reveal that the acidic groups of the glycans are key specificity determinants for enzymes and binding proteins that make up the degradative apparatus.

Author contributions: A.C., E.C.L., H.J.G., and D.N.B. designed research; A.C., E.C.L., A.B., S.J.F., D.A.N., H.M., N.T., V.L., and M.C. performed research; J.E.T. contributed new reagents/analytic tools; A.C., E.C.L., A.B., S.J.F., H.M., N.T., V.L., B.H., M.C., and D.N.B. analyzed data; and A.C., E.C.L., B.H., J.E.T., M.C., H.J.G., and D.N.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The crystallography, atomic coordinates, and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4AK1, 4AK2, 5G2T, 5G2U, and 5G2V).

[1]A.C. and E.C.L. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: david.bolam@ncl.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1704367114/-/DCSupplemental.

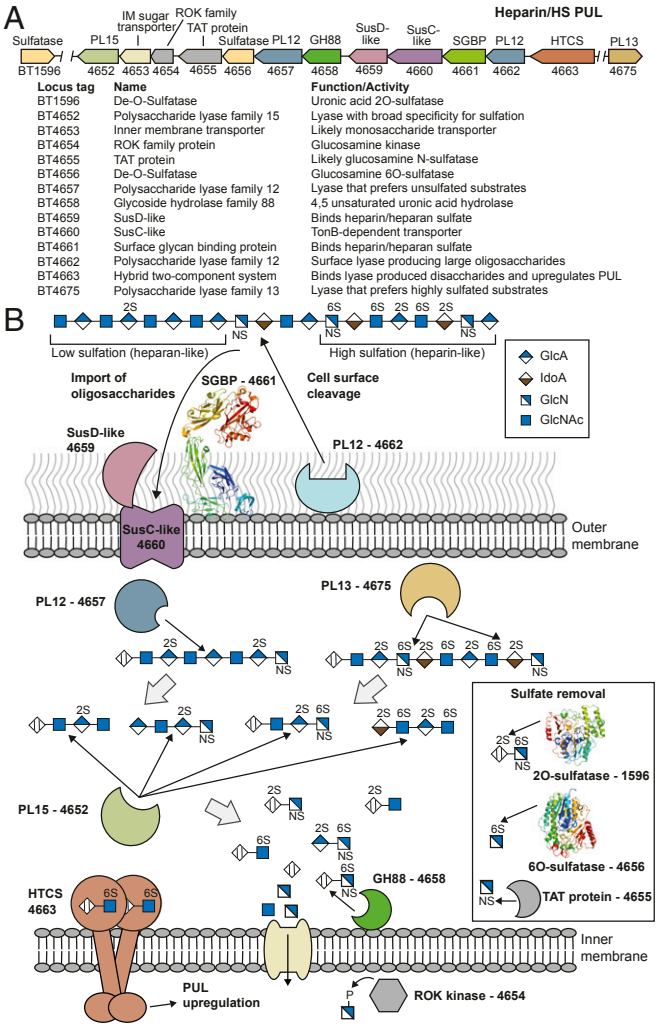

Fig. 1. Model for degradation of Hep and HS by *Bt*. (*A*) Schematic of the PUL_{Hep} locus in *Bt*. IM, inner membrane. (*B*) Schematic representation of the cellular location, activity, and specificity of the PUL_{Hep}-encoded enzymes and glycan binding proteins. A typical HS structure is shown. The proteins for which X-ray crystallographic structures were determined in this work are represented by thumbnail images of the structure.

Hep I–III. Hep I lyases require sulfation, Hep II PLs exhibit promiscuity with respect to sulfation patterns, and Hep III enzymes cleave low-sulfation regions of these GAGs (13, 14). The lyases generate products capped by Δ4,5-unsaturated UA, which is released from the GAG by GHs from family GH88 (13). In addition, some GAG-specific sulfatases have been characterized (15).

Although enzymes active against Hep and HS have been described, how these PLs, GHs, and sulfatases are tailored to act in unison to address the sulfation problem posed by these GAGs is unclear. Several scenarios can be proposed, including desulfation before backbone cleavage, the exploitation of broad specificity PLs that can cleave Hep and HS independent of their sulfation, or the recruitment of a consortium of lyases, with each enzyme targeting specific substructures in these GAGs that differ in their sulfation profiles. Similarly, the mechanisms by which the cell surface glycan binding proteins (SGBPs), which contribute to the glycan degradation machinery, target conserved features of these heterogeneous glycans are unknown.

Bacteroides thetaiotaomicron (*Bt*), a member of the CM, uses Hep and HS as high-priority nutrient sources that activate a single PUL (PUL_{Hep}) (16). Here, we have dissected the mechanism by which the enzymes and binding proteins encoded by this locus fully deconstruct these highly variable GAGs and thus, solve the sulfation problem. Crystal structures of key proteins provide mechanistic insight into substrate and ligand recognition. The data revealed how the specificity of the apparatus is optimized to target the repertoire of GAG structures presented to the bacterium at the different stages of the degradative process.

Results

Structure of HS and Hep. Hep and HS are composed of a disaccharide repeating unit comprising a UA, D-glucuronic (GlcA), or L-iduronic acid (IdoA) alternating with D-glucosamine (GlcN). GlcN is linked α1,4 to the UA, whereas IdoA and GlcA are linked α1,4 and β1,4, respectively, to the amino sugar (10) (Fig. 1). GlcN can be sulfated on O6 and sulfated or acetylated at N2 (GlcNAc), whereas the UAs are often sulfated on O2. Hep contains significantly more IdoA than HS and is almost completely sulfated, whereas HS varies in its sulfation pattern, containing highly sulfated regions (Hep-like) and areas that contain little or no sulfation (Fig. 1). The DP of Hep (~40) is substantially less than that of HS (~80–200).

PUL_{Hep} Structure. *Bt* grows on Hep, HS, and *N*-acetyl-de–*O*-sulfated heparin (ΔSHep) (Fig. S1) (16, 17). Transcriptomics of *Bt* cultured on Hep indicate that only PUL_{Hep}, extending from *bt4652* to *bt4675*, was up-regulated in response to this GAG (16). The locus encodes PLs, sulfatases, a GH88, and a member of the repressor, ORF, kinase (ROK; Pfam PF00480) family (Fig. 1). In addition, the sulfatase BT1596 is also up-regulated by Hep (15). PUL_{Hep} encodes a single SusC/SusD-like outer membrane glycan transporter system and a potential SGBP (Fig. 1). In silico analysis of the occurrence of PUL_{Hep} suggests that it is widely distributed within the *Bacteroides* and retains high sequence conservation and synteny (Fig. S2).

SGBPs. The extracellular location of the putative SGBP, BT4661, was revealed by immunofluorescence and proteinase K treatment using antibodies against the protein (Fig. S3 *A* and *B*). Immunoprecipitation of BT4661 from Hep-grown cells followed by Western blotting revealed the presence of $BT4659^{SusD\text{-}like}$, indicating that the two proteins physically associate (Fig. S3*C*). The interaction between the PUL-encoded SGBP and SusD-like is consistent with previous data from the starch utilization system of *Bt* (18). The importance of $BT4659^{SusD\text{-}like}$ in Hep/HS utilization was highlighted by the severe lag (>20 h) displayed by $\Delta bt4659^{SusD\text{-}like}$ on these GAGs (Fig. S1*A*). By contrast, $\Delta bt4661^{SGBP}$ had no growth defect on HS or ΔSHep but had a noticeable phenotype when grown on Hep-derived oligosaccharides (Fig. S1*A*, *Inset*). These data suggest that the SGBP plays a role in oligosaccharide scavenging rather than polysaccharide acquisition, such as observed in *Bacteroides ovatus*-mediated xyloglucan degradation (19).

Glycan specificity. Isothermal titration calorimetry (ITC) (Fig. 2*A* and Table S1) revealed that $BT4661^{SGBP}$ bound tightly to both Hep and HS (K_d ~ 3 μM) but did not bind to other GAGs. Notably, coverage on both GAGs was similar, suggesting the protein can tolerate both sulfated and unsulfated regions and could interact or accommodate the different UAs and neutral sugars in Hep and HS (Fig. 1). $BT4661^{SGBP}$ bound to oligosaccharides with a DP ≥ 4, with affinity increasing up to DP = 6 to a level similar to Hep. These data indicate that the $BT4661^{SGBP}$ binding site can accommodate a hexasaccharide and supports the ability of the SGBP to tolerate/recognize sulfation and IdoA at all subsites (Table S1). These data show that $BT4661^{SGBP}$ is an SGBP of the Hep/HS degrading apparatus and that the protein can accommodate a range of Hep-based GAG structures.

$BT4659^{SusD\text{-}like}$ bound Hep, HS, and ΔSHep with similar affinity, but the K_d was ~20- to 30-fold lower than that for the $BT4661^{SGBP}$. Differences in affinity between the SusD-like and

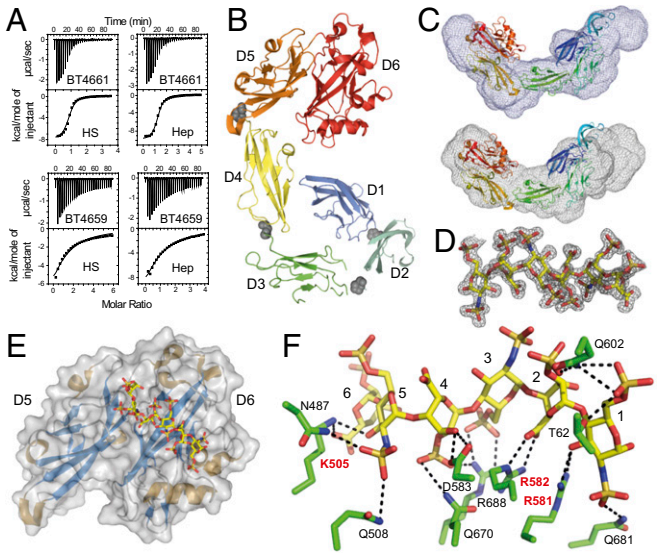

Fig. 2. Structural and biochemical characterization of the PUL_{Hep}-encoded SGBP BT4661. (*A*) ITC traces for (*Upper*) $BT4661^{SGBP}$ and (*Lower*) $BT4659^{SusD\text{-}like}$ against HS and Hep. (*B*) Cartoon representation of $BT4661^{SGBP}$ colored from blue to red from the N terminus to the C terminus. Each of six discrete domains is labeled. Gray spheres show the positions of the interdomain prolines. (*C*) SAXS envelopes (gray mesh) for (*Upper*) apoBT4661^{SGBP} and (*Lower*) $BT4661^{SGBP}$ + Hep. The best fit of the crystal structure of $BT4661^{SGBP}$ (colored from blue to red from the N terminus to the C terminus) is shown inside the SAXS envelopes. (*D*) 2Fo-Fc map contoured at 1.0 sigma for fully sulfated Hep-derived hexasaccharide in complex with $BT4661^{SGBP}$. (*E*) Structure of D5 and D6 C-terminal domains of $BT4661^{SGBP}$ bound to Hep hexasaccharide (carbons as yellow sticks). (*F*) Ligand binding site of $BT4661^{SGBP}$ bound to fully sulfated Hep hexasaccharide. Amino acid side chains are shown in green, the sugars are in yellow, and H bonds between the ligand and protein are black dotted lines. The residues labeled in red are critical for ligand recognition. The six sugar binding subsites are labeled one to six from the reducing end of the oligosaccharide.

cognate SGBP have been reported previously and may reflect the proposed different roles that these proteins play in glycan utilization (5, 19).

Structure of full-length BT4661^{SGBP}. The crystal structure of $BT4661^{SGBP}$ (Table S2) revealed six discrete domains that adopt an Ig-like fold (Fig. 2*B*). This multi–Ig-like domain structure is common to the other three SGBPs, which have structures that are known, although little or no sequence identity is evident between these proteins (5, 19, 20). The domains, defined as D1 (N terminus) to D6 (C terminus), are arranged in an extended, curved conformation. Interdomain proline residues may limit the flexibility of the SGBP (19). Small angle X-ray scattering (SAXS) on full-length $BT4661^{SGBP}$ in the presence and absence of Hep gave the same R_g and D_{max} values for both conditions (Table S3), indicating that no large conformational changes are imposed by the target GAG (Fig. 2*C*, *SI Results*, and Fig. S3*D*).

Structure of truncated BT4661^{SGBP} bound to oligosaccharide. Attempts to obtain a ligand complex of full-length BT4661 were unsuccessful. A truncated form of the protein comprising only D5 and D6 displayed similar ligand binding to the WT SGBP. The crystal structure of the D5/6 derivative ($TrBT4661^{SGBP}$) was determined in complex with a fully sulfated, Hep-derived unsaturated hexasaccharide (Δ4,5UA2S-GlcNS6S-IdoA2S-GlcN6S-IdoA2S-GlcNS6S) (Fig. 2 *D–F* and Table S2). The binding site is formed across the D5 and D6 domains, with the reducing end of the hexasaccharide, GlcNS6S, lying in D6 (subsite 1) and the nonreducing end ΔUA2S located in D5 (subsite 6). D6 and D5 house sugars 1–4 and 5 and 6 of the hexasaccharide, respectively. The presence of the glycan binding site on the C-terminal domain of an extended multidomain SGBP has been observed previously and may be an adaptation to enable ligand binding to occur at a distance from the cell surface (19). Similar to other GAG binding proteins, interactions between ligand and protein mainly involved basic amino acids rather than aromatic residues, which is the characteristic signature of protein–carbohydrate recognition of neutral glycans (10, 19–21). Alanine scanning mutagenesis revealed that K505, R581, and R582 are the key residues involved in ligand binding (Fig. S3*E*). R582 and K505 interact with the carboxylates of the UA at subsites 2 and 6, respectively, whereas R581 interacts with O3 of the reducing end sugar. These interactions will be conserved in all Hep-based GAGs, providing an explanation for the ability of BT4661 to recognize both HS and Hep with similar affinity.

Enzymes Encoded by PUL_{Hep}. The activity of the PLs sulfatases and GH encoded by PUL_{Hep} was determined against different substrates, and the growth of mutant strains lacking active forms of these enzymes on Hep, HS, and ΔSHep was assessed (Fig. S1 *B–F*).

Surface lyase BT4662^{PL12}. The lipoprotein $BT4662^{PL12}$ was shown to be a surface enzyme (Fig. S3*B*) that was active on Hep, ΔSHep, and HS, with a preference for the latter substrate (Fig. 3*A* and Table S4). Aerobic whole-cell assays, which report only on the activity of surface enzymes, showed that the pattern of reaction products closely matches that of recombinant $BT4662^{PL12}$ (Fig. S4*A*), confirming the cellular location of the PL. The range of oligosaccharides generated during the early stages of GAG degradation points to an endo-activity (Fig. 3*A* and Fig. S4*E*). In other characterized PULs, a surface endo-acting glycanase plays a key role in polysaccharide utilization by generating oligosaccharides that are of an appropriate size to be transported by the SusC/D-like apparatus (5). However, the Δ*bt4662^{PL12} Bt* mutant displayed only a partial growth defect on HS (Fig. S1*B*), likely reflecting the heterogeneity of the DP of the GAG; thus, the smaller forms of this glycan were able to be transported into the periplasm of *Bt* without the need for prior depolymerization. Growth on Hep and ΔSHep was not impaired in the Δ*bt4662^{PL12}* mutant, reflecting their low DP, which likely enables these molecules to be transported into the periplasm without prior enzymatic processing.

End point assays revealed that Hep was almost completely inaccessible to $BT4662^{PL12}$ (only ∼5% degradation), whereas the depolymerization values of HS and ΔSHep were ∼40 and ∼80%, respectively (Table S5). Against HS, the limit products comprised a wide range of products, whereas complete degradation of ΔSHep

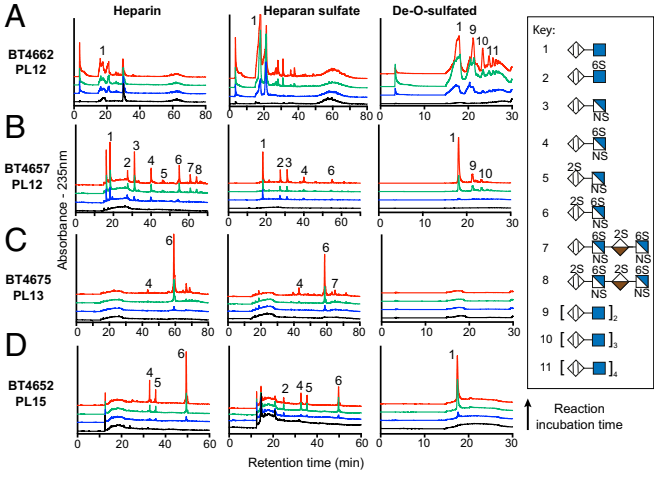

Fig. 3. Product profile of the PUL_{Hep}-encoded PLs against Hep, HS, and ΔSHep. (*A*) BT4662, (*B*) BT4657, (*C*) BT4675, and (*D*) BT4652. Peaks were identified by comparison with known standards at A_{235}. Black lines represent zero time points, whereas blue, green, and red lines represent early, middle, and late points, respectively, in the reaction time course for each enzyme.

generated oligosaccharides with a DP of 2–10 (Fig. 3*A* and Fig. S4*E*). BT4662^{PL12} was only active against Hep oligosaccharides with a DP ≥ 10 (Table S4). These data suggest BT4662^{PL12} has a large substrate binding cleft, in which sulfate groups can only be accommodated at specific positions, and the enzyme displays an endo mode of action, with limited processivity. These data are consistent with a role for the enzyme in depolymerizing high molecular weight GAGs, such as HS, at the bacterial surface.

Periplasmic PLs. Based on proteinase K treatment and the product profiles of whole-cell assays compared with those of the recombinant enzymes, the majority of the polysaccharide-degrading enzymes encoded by PUL$_{Hep}$ are located in the periplasm (Figs. S3*B* and S4*A*) and display no evidence for metal dependence. A unifying feature of the three periplasmic PLs, BT4652^{PL15}, BT4657^{PL12}, and BT4675^{PL13}, is the appearance of limit disaccharide products during the initial phase of degradation (Fig. 3 *B–D*). This activity is indicative of a degree of processivity and likely results in rapid production of the disaccharides that act as the signaling cue(s) required to up-regulate PUL$_{hep}$ (17). The three periplasmic lyases have differing preferences for Hep, HS, and ΔSHep (described in detail below), reflecting the ability of *Bt* to use these three substrates.

BT4657^{PL12} exhibited the greatest activity against HS and could access ∼50% of the polymer; while against Hep, the lyase cleaved 27% of the glycosidic bonds (Tables S4 and S5). The enzyme completely degraded ΔSHep to a single unsulfated disaccharide species (Fig. 3*B* and Tables S4 and S5). During the initial degradation of ΔSHep, BT4657^{PL12} generated a range of oligosaccharides, with the disaccharide being the dominant product (Fig. 3*B* and Fig. S4*E*). These data indicate that BT4657^{PL12} is primarily endo-acting, with a degree of processivity. The products generated from partially sulfated oligosaccharides showed that BT4657^{PL12} is unable to accommodate 2-*O* sulfation of UA in its active site (+1 subsite), explaining why the enzyme displayed limited activity against Hep (*SI Results*).

In contrast, BT4675^{PL13} displayed highest activity against Hep and was inactive against ΔSHep, consistent with a previous study suggesting that sulfation was required for activity (Table S4) (22). The requirement for sulfation was mirrored in the end point assays, which showed almost complete digestion of Hep but only ∼50% of HS (Table S5). The PL13 enzyme generated a range of different size products indicative of an endo mode of action but also displayed significant processivity, because UA2S-GlcNS6S was the dominant product early in the reaction (Fig. 3*C*).

PL15 enzymes have previously only been shown to be alginate lyases (23). BT4652^{PL15}, however, was active on all three GAGs tested. The enzyme was ∼800-fold more active against fully sulfated Hep oligosaccharides compared with ΔSHep, indicating that sulfation was key for optimal activity (Table S4), consistent with the production of UA2S-GlcNS6S, UA2S-GlcNS, and UA-GlcNS6S from Hep; however, against HS, an additional disaccharide, UA-GlcNAc6S, was generated, and the unsulfated disaccharide UA-GlcNAc was produced only from ΔSHep (Fig. 3*D*). BT4652^{PL15} generated disaccharides against all substrates; no intermediate products were observed during initial degradation (Fig. 3*D* and Fig. S4*D*). These data indicate the enzyme is exclusively exo-processive and can accommodate sulfates at all positions of the substrate. Although sulfation is not essential, it is a specificity determinant, and as such, BT4652^{PL15} displays more substrate flexibility than BT4675^{PL13}.

The data described above suggest that BT4652^{PL15} and BT4675^{PL13} can target similar heavily *O*-sulfated regions of Hep and HS, although the PL15 enzyme can, in addition, cleave sparsely sulfated regions of these GAGs. Consistent with this view is the observation that mutant strains lacking BT4652^{PL15} or BT4675^{PL13} displayed growth defects on Hep and to a lesser extent, HS but not ΔSHep (Fig. S1 *B–E*). The growth properties of Δ*bt4657*^{PL12} revealed a significant role for BT4657^{PL12} in HS degradation. The endo-acting lyase likely cleaves areas low in sulfation, creating nonreducing ends that are targeted by BT4652^{PL15} and BT4675^{PL13}. Our data thus provide a biological context for the multiple lyases expressed by *Bt* in response to Hep and HS (*Discussion*). The primary substrates for BT4675^{PL13} and BT4657^{PL12} are regions of the GAGs that were highly and poorly sulfated, respectively. The Δ*bt4652*^{PL15} strain had the largest growth defect on Hep, whereas Δ*bt4652*^{PL15}/Δ*bt4657*^{PL12} displayed little growth on HS and a marked increase in the lag phase when cultured on ΔSHep (Fig. S1*D*). These growth profiles suggest that BT4657^{PL12} and BT4675^{PL13} produce oligosaccharides that only BT4652^{PL15} can degrade, and thus, the additional flexibility in recognition by the PL15 bridges the two complementary activities of the PL12 and PL13 enzymes (discussed in detail below).

GH88 enzyme. BT4658^{GH88} belongs to a family of enzymes that cleave the glycosidic linkage between the Δ4,5-unsaturated UA and GlcN/GlcNAc disaccharides that are produced by GAG lyases. The enzyme was not active when O2 of the UA was sulfated but could accommodate sulfation at N2 or O6 of the neutral sugar, mirroring the binding specificity of the PUL$_{Hep}$ hybrid two-component system (HTCS) BT4663 (Table S6) (17). The tuning of the specificity of the GH88 enzyme with that of its cognate HTCS is analogous to that observed in the *Bt* chondroitin sulfate PUL, indicating a conserved role for these enzymes in controlling the rate of signal degradation during growth on GAGs (24). Although BT4658^{GH88} displayed similar catalytic efficiencies against GlcNS and GlcNAc, the K_m and k_{cat} for the sulfated sugar were lower, suggesting that the *N*-linked sulfate contributes to substrate binding but restricts product departure, leading to a reduced k_{cat}.

O-sulfatases. A previous study showed that BT1596$^{2S\text{-}sulf}$ and BT4656$^{6S\text{-}sulf}$ are exo-acting *O*-sulfatases; BT1596$^{2S\text{-}sulf}$ targets the O2 sulfation of unsaturated UA at the nonreducing end of di- and oligosaccharides, whereas BT4656$^{6S\text{-}sulf}$ cleaves the O6-sulfate of the monosaccharide GlcNAc6S or GlcNS6S but is inactive against GalNAc6S (Table S6) (15). To explore the mechanism of substrate recognition, the crystal structures of inactive forms of the two *Bt O*-sulfatases were determined in complex with their substrates.

BT1596$^{2S\text{-}sulf}$ and BT4656$^{6S\text{-}sulf}$ share a conserved α/β hydrolase fold comprising a single domain (Fig. 4, *Upper* and Table S2). Both enzymes have a pocket topology with metal binding sites located at the base of the pocket. Calcium was modeled into the BT4656$^{6S\text{-}sulf}$ ligand-bound structure, whereas no metal was observed in BT1596$^{2S\text{-}sulf}$ (Fig. 4, *Lower*). Alanine substitution of the residues in this pocket inactivated both enzymes, revealing the active site location (*SI Results* and Table S6).

Inactive forms of BT1596$^{2S\text{-}sulf}$ and BT4656$^{6S\text{-}sulf}$ were cocrystallized with ΔUA2S-GlcNS6S and GlcNS6S, respectively (Fig. 4). BT1596$^{2S\text{-}sulf}$ and BT4656$^{6S\text{-}sulf}$ were inactive when expressed in *Escherichia coli*, because the enteric bacterium cannot convert S64 and S77, respectively, into formylglycine, which functions as the catalytic nucleophile [to generate the formylglycine in *E. coli*, the catalytic serine was mutated to cysteine because the bacterium can convert cysteine to formylglycine (25)]. The essential histidines H188 in BT1596$^{2S\text{-}sulf}$ and H196 in BT4656$^{6S\text{-}sulf}$ are in close proximity to the scissile bond (Fig. 4, *Lower* and Table S6). Thus, H188 and H196 are ideal candidates for the catalytic acid that is required to protonate the O2/O6 of the departing sugar after nucleophilic attack by the formylglycine and formation of a sulfate–enzyme intermediate, which is then cleaved via a β-elimination to complete the catalytic cycle.

In BT1596$^{2S\text{-}sulf}$, the carboxylate of ΔUA2S makes a bidentate ionic interaction with R237, and loss of this interaction (R237A) results in a 2,000-fold reduction in catalytic efficiency (Fig. 4*A* and Table S6). This interaction is likely the major specificity determinant used by the enzyme and explains why the sulfatase acts on chondroitin disaccharides, which also possess O2-sulfated UA. In BT4656$^{6S\text{-}sulf}$, R363 and D361 coordinate O4, and mutation of R363 to Ala caused an ∼500-fold decrease in activity, whereas

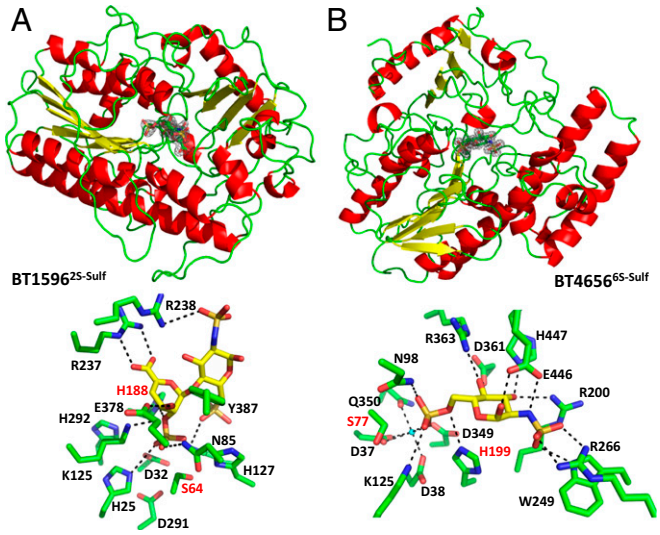

Fig. 4. Sulfatase structures. *Upper* shows cartoon representations of (*A*) BT1596$^{2S\text{-}sulf}$ and (*B*) BT4656$^{6S\text{-}sulf}$. 2Fo-Fc maps contoured at 1.0 sigma for Δ4,5UA2Sβ1–4GlcNS6S in complex with BT1596$^{2S\text{-}sulf}$ and GlcNS6S in complex with BT4656$^{6S\text{-}sulf}$ are shown in the respective enzymes active sites. *Lower* shows stick representations of the active site interactions between (*A*) BT1596$^{2S\text{-}sulf}$ and Δ4,5UA2Sβ1–4GlcNS6S and (*B*) BT4656$^{6S\text{-}sulf}$ and GlcNS6S. Amino acid side chains are shown in green, the sugars are in yellow, and H bonds between the sugar and protein are black dotted lines. Residues labeled in red are the key catalytic amino acids.

D361A was inactive (Table S6). These amino acids are likely specificity determinants for gluco- over galacto-configured substrates. (*SI Results* has a description of all mutants.)

N-*sulfatase*. There are no predicted sulfamidases in PUL$_{Hep}$; however, GlcNS does not accumulate during growth on Hep, indicating that *Bt* uses this sugar (Fig. S4*F*). We deleted the remaining uncharacterized ORF in PUL$_{Hep}$, *bt4655*. This mutant displayed a growth defect on sulfated GAGs, and GlcNS accumulated in the media during growth on Hep (Figs. S1*F* and S4*F*). These data indicate that BT4655 cleaves the sulfamate linkage and is a previously uncharacterized class of sulfatase. Unfortunately, the proposed role of BT4655 as a sulfamidase could not be confirmed biochemically, because a soluble recombinant form of the protein could not be generated.

The data show that the three sulfatases are key for both Hep and HS breakdown. BT1596$^{2S\text{-}sulf}$ is the first to act, removing 2-O sulfation from the limit PL-derived disaccharide and allowing BT4658^{GH88} to hydrolyze its target linkage, because this enzyme does not use 2-O–sulfated substrates. The sulfated glucosamine monosaccharides are then substrates for BT4656$^{6S\text{-}sulf}$ and the likely sulfamidase BT4655 (Table S6).

Cytoplasmic Sugar Kinase. BT4654^{ROK} is a cytoplasmic member of the ROK protein family that possesses the conserved DxGxT motif and thus, likely an ATP-dependent kinase (26). This activity was confirmed by the capacity of the enzyme to phosphorylate gluco- and manno-configured substrates but not galacto-configured sugars (Table S6). The k_{cat}/K_m of the enzyme decreased from GlcN > GlcNAc >> Glc/Man >> GlcNS, which was driven by increases in K_m (Table S6). These data show that O4 is critical for catalysis, whereas strong selection in affinity is made at C2 for an equatorial amine, indicating that BT4654^{ROK} uses both an equatorial N2 and an O4 as specificity determinants. Because the Δ*bt4654*^{ROK} mutant displayed no growth defect on Hep or HS (Fig. S1*F*), *Bt* seems to display redundancy in GlcN and GlcNAc phosphorylation.

Discussion

Bt PUL$_{Hep}$ orchestrates the hierarchical degradation of Hep and HS. Both of these GAGs could be derived from dietary meat or host glycans. In vivo expression of PUL$_{Hep}$ in mice colonized with *Bt* has only been observed in bacteria occupying the mucosal layer, suggesting that the source of Hep/HS available to *Bt* is from the turnover of epithelial cells rather than the diet (27). HS could also be supplied in the human diet as a component of meat; however, colonic metagenomic data obtained in a study, which swapped subjects from vegetarian diets to meat-rich diets, did not reveal any PL families associated with the degradation of Hep/HS (28). These data support the hypothesis that the source of Hep/HS in the human gut accessible to the microbiota is from the turnover of epithelial cells. Goodman et al. (29) identified through transposon mutagenesis genes in *Bt* important for colonization of the mouse gut. Strains with insertions in the *bt4659*$^{susD\text{-}like}$ were significantly decreased in the output from the monocolonized mouse gut compared with input. Interestingly, in mice cocolonized with six other Bacteroidetes strains in addition to the *Bt* mutant population, *bt4658*^{GH88} and *bt4661*^{SGBP} insertion mutants were much lower in abundance than in monocolonized mice, indicating that the ability to degrade HS is under increased selection pressure for *Bt* in the presence of other Bacteroidetes (Fig. S2).

Within the Hep-based GAGs, there is substantial structural diversity, especially in terms of the level of sulfation, and it is this heterogeneity that provides an explanation for the distinct but complementary activities of the enzymes encoded by the PUL$_{Hep}$. A key feature of the PUL$_{Hep}$ degradative apparatus is the surface lyase, BT4662^{PL12}, which although dispensable for growth on Hep, plays a significant role in optimal HS utilization. This difference likely reflects the high DP of HS relative to Hep, and thus, the GAG needs to be subjected to a degree of extracellular depolymerization to generate molecules that can be imported into the periplasm (5).

In the periplasm, the substrate specificity of endoprocessive lyases is optimized to target highly sulfated (BT4675^{PL13}) or low/unsulfated regions (BT4657^{PL12}), producing small oligosaccharides that only the exoprocessive lyase (BT4652^{PL15}) is able to degrade. This role for BT4652^{PL15} is critical, because loss of the enzyme means that elements of both Hep and HS remain inaccessible to *Bt*. The presence of a PL15 lyase is a conserved feature of Hep/HS utilization in other Bacteroidetes, supporting its key role in breakdown of these GAGs (Fig. S2). The synergistic specificities displayed by the PUL$_{Hep}$ lyases enable the bacterium to cleave the backbone of GAGs that display substantial structural variation, particularly in their sulfation patterns, leading to the generation of disaccharides. The disaccharides are broken down by the complementary activities of the sulfatases and the GH88 enzyme-generating sugars that can then be metabolized by the bacterium.

Bt places a high priority on Hep and HS utilization shown by the lack of repression of PUL$_{Hep}$ by other polysaccharides and glucose (30). HS is an abundant source of GlcNAc, which is required for the synthesis of peptidoglycan. Although *Bt* seems to contain all of the biosynthetic genes needed to make GlcNAc, it may still prioritize GlcNAc derived from glycans to synthesize its cell wall. As stated above, we propose that *Bt* uses HS/Hep from the host mucosa rather than the diet. Because the epithelium is likely to contain much higher amounts of HS compared with Hep, we believe that the biologically relevant glycan targeted by PUL$_{Hep}$ is HS. Analysis of the genomes of other members of the CM revealed PULs similar to *Bt* PUL$_{Hep}$ (Fig. S2). *Bacteroides xylanisolvens* and *Bacteroides ovatus* both contain a predicted surface PL12 lyase similar to BT4662^{PL12}, suggesting that these organisms also target HS. In contrast, the GAG-degrading apparatus of *Bacteroides intestinalis* lacks an obvious surface Hep/HS lyase (Fig. S2). Thus, the bacterium may target oligosaccharides or low molecular weight

BIOCHEMISTRY

GAGs, such as Hep, which can be imported without the need for enzymatic depolymerization.

The structural data of the sulfatases showed that the interactions with the target sulfate are highly conserved in BT1596$^{2S\text{-}sulf}$, BT4656$^{6S\text{-}sulf}$, and the human 6S GalNAc sulfatase (GALNSase) (Fig. S3*F*). In BT4656$^{6S\text{-}sulf}$ and GALNSase (the only other 6-O carbohydrate sulfatase for which a structure is available; Protein Data Bank ID code 4FDI), however, there is very little conservation in the residues that interact with the carbohydrate component of the substrate (Fig. S3*G*), showing that this area is highly variable and could be the basis for bespoke highly specific glycosulfatase inhibitors. Such inhibitors may have therapeutic relevance, because a recent study showed that, in a colitis-sensitive mouse model, sulfatases expressed by *Bt* are essential factors in inducing the disease (31).

Conclusions

This report describes the characterization of a PUL that orchestrates the degradation of Hep/HS, high-priority heterogeneous glycans. Unlike other microbial systems that remove side chains before backbone degradation, the glycan backbone of Hep/HS is depolymerized before removal of the sulfate groups. To achieve this depolymerization, *Bt* expresses lyases with complementary activities that, in combination, can fully deconstruct the Hep-based GAGs. The PLs, sulfatases, and GH encoded by PUL$_{Hep}$ provide an example of how an enzyme consortium can degrade substrates that display highly variable sulfation patterns. Understanding the mechanisms by which gut bacteria depolymerize GAGs may provide crucial insights into how the human gut microbiota is adapted to use nonmucin host glycans and the role that this process plays in colonization and survival in the gut. The model developed here can be harnessed to interrogate metagenomic data to explore the mechanisms by which different members of the microbiota metabolize these complex GAGs.

Materials and Methods

***Bacteroides* Culture and Genetic Manipulation.** *Bt* VPI-5482 was cultured anaerobically, and genomic null mutations were generated as described previously (16, 32).

Enzyme Assays. PL activity was monitored at A_{235}. BT4658^{GH88} glucuronyl hydrolase activity was determined by monitoring loss of signal at A_{235}. Sulfatase and kinase activities were determined through either linked assays or HPLC. Products were analyzed by TLC and/or HPLC.

Crystallization of BT1596, BT4656, and BT4661. Crystallization and structure determination were as described in *SI Materials and Methods*.

Full details of all experimental procedures used are described in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Carl Morland for expert technical assistance. This work was funded by the United Kingdom Biotechnology and Biological Sciences Research Council (Grant BB/F014163/1), the European Research Council (Grant 322820), and the Wellcome Trust (Grant WT097907MA).

1. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI (2011) Human nutrition, the gut microbiome and the immune system. *Nature* 474:327–336.
2. Round JL, Mazmanian SK (2009) The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 9:313–323.
3. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI (2005) Host-bacterial mutualism in the human intestine. *Science* 307:1915–1920.
4. Koropatkin NM, Cameron EA, Martens EC (2012) How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* 10:323–335.
5. Rogowski A, et al. (2015) Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun* 6:7481.
6. Ndeh D, et al. (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* 544:65–70.
7. Martens EC, Koropatkin NM, Smith TJ, Gordon JI (2009) Complex glycan catabolism by the human gut microbiota: The Bacteroidetes Sus-like paradigm. *J Biol Chem* 284:24673–24677.
8. Cao Y, Rocha ER, Smith CJ (2014) Efficient utilization of complex N-linked glycans is a selective advantage for Bacteroides fragilis in extraintestinal infections. *Proc Natl Acad Sci USA* 111:12901–12906.
9. Tailford LE, et al. (2015) Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nat Commun* 6:7624.
10. Gandhi NS, Mancera RL (2008) The structure of glycosaminoglycans and their interactions with proteins. *Chem Biol Drug Des* 72:455–482.
11. Oshiro M, et al. (2001) Immunohistochemical localization of heparan sulfate proteoglycan in human gastrointestinal tract. *Histochem Cell Biol* 115:373–380.
12. Lombard V, et al. (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem J* 432:437–444.
13. Nakamichi Y, Mikami B, Murata K, Hashimoto W (2014) Crystal structure of a bacterial unsaturated glucuronyl hydrolase with specificity for heparin. *J Biol Chem* 289:4787–4797.
14. Ulaganathan T, et al. (2017) Conformational flexibility of PL12 family heparinases: Structure and substrate specificity of heparinase III from Bacteroides thetaiotaomicron (BT4657). *Glycobiology* 27:176–187.
15. Ulmer JE, et al. (2014) Characterization of glycosaminoglycan (GAG) sulfatases from the human gut symbiont Bacteroides thetaiotaomicron reveals the first GAG-specific bacterial endosulfatase. *J Biol Chem* 289:24289–24303.
16. Martens EC, Chiang HC, Gordon JI (2008) Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* 4:447–457.
17. Lowe EC, Baslé A, Czjzek M, Firbank SJ, Bolam DN (2012) A scissor blade-like closing mechanism implicated in transmembrane signaling in a Bacteroides hybrid two-component system. *Proc Natl Acad Sci USA* 109:7298–7303.
18. Cho KH, Salyers AA (2001) Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* 183:7224–7230.
19. Tauzin AS, et al. (2016) Molecular dissection of xyloglucan recognition in a prominent human gut symbiont. *MBio* 7:e02134–e15.
20. Cameron EA, et al. (2012) Multidomain carbohydrate-binding proteins involved in Bacteroides thetaiotaomicron starch metabolism. *J Biol Chem* 287:34614–34625.
21. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem J* 382:769–781.
22. Han YH, et al. (2009) Structural snapshots of heparin depolymerization by heparin lyase I. *J Biol Chem* 284:34019–34027.
23. Ochiai A, Yamasaki M, Mikami B, Hashimoto W, Murata K (2010) Crystal structure of exotype alginate lyase Atu3025 from Agrobacterium tumefaciens. *J Biol Chem* 285:24519–24528.
24. Raghavan V, Lowe EC, Townsend GE, 2nd, Bolam DN, Groisman EA (2014) Tuning transcription of nutrient utilization genes to catabolic rate promotes growth in a gut bacterium. *Mol Microbiol* 93:1010–1025.
25. Dierks T, et al. (1998) Posttranslational formation of formylglycine in prokaryotic sulfatases by modification of either cysteine or serine. *J Biol Chem* 273:25560–25564.
26. Titgemeyer F, Reizer J, Reizer A, Saier MH, Jr (1994) Evolutionary relationships between sugar kinases and transcriptional repressors in bacteria. *Microbiology* 140:2349–2354.
27. Li H, et al. (2015) The outer mucus layer hosts a distinct intestinal microbial niche. *Nat Commun* 6:8292.
28. David LA, et al. (2014) Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505:559–563.
29. Goodman AL, et al. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6:279–289.
30. Rogers TE, et al. (2013) Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol Microbiol* 88:876–890.
31. Hickey CA, et al. (2015) Colitogenic Bacteroides thetaiotaomicron antigens access host immune cells in a sulfatase-dependent manner via outer membrane vesicles. *Cell Host Microbe* 17:672–680.
32. Koropatkin NM, Martens EC, Gordon JI, Smith TJ (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* 16:1105–1115.
33. Reith J, Berking A, Mayer C (2011) Characterization of an N-acetylmuramic acid/N-acetylglucosamine kinase of Clostridium acetobutylicum. *J Bacteriol* 193:5386–5392.
34. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235–242.
35. Terrapon N, Lombard V, Gilbert HJ, Henrissat B (2015) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* 31:647–655.
36. Terrapon N, Weiner J, Grath S, Moore AD, Bornberg-Bauer E (2014) Rapid similarity search of proteins using alignments of domain arrangements. *Bioinformatics* 30:274–281.
37. Guinier A, Fournet F (1955) *Small Angle Scattering of X-Rays* (Wiley Interscience, New York).
38. Svergun D (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25:495–503.
39. Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80:2946–2953.
40. Petoukhov MV, Eady NA, Brown KA, Svergun DI (2002) Addition of missing loops and domains to protein models by x-ray solution scattering. *Biophys J* 83:3113–3125.
41. Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36:860–864.
42. Davies GJ, Wilson KS, Henrissat B (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* 321:557–559.

Supporting Information

Cartmell et al. 10.1073/pnas.1704367114

SI Materials and Methods

Sources of Carbohydrates. Carbohydrates were from Dextra Laboratories, except for HS, which was either from Iduron or a gift from J. Turnball (University of Liverpool, Liverpool, United Kingdom).

***Bacteroides* Culture and Genetic Manipulation.** *Bt* VPI-5482 was cultured anaerobically in a Whitley A35 Workstation (Don Whitley) at 37 °C in either tryptone yeast extract glucose medium or minimal medium (MM) as described previously (5, 7). *Bt* strains containing specific gene deletions or inactivated versions of enzymes were made by counterselectable allelic exchange as described previously (32). The $\Delta BT4654^{ROK}$, $\Delta BT4655^{NS\text{-}Sulf}$, $\Delta BT4675^{PL13}$, and $\Delta BT4662^{PL12}$ strains are all deletion mutants, whereas for the $\Delta BT4652^{PL15}$ and $\Delta 4657^{PL12}$ strains, only the key active site residues were mutated in the respective enzymes. Growth of the WT and mutants was measured on glycans by mixing the autoclave-sterilized polysaccharides (1% final) with MM and monitoring growth continuously in 96-well plates using a Biotek Epoch plate reader. Growth curves presented are averages of three biological replicates.

Microscopy. Glucose-grown cells were harvested at midexponential phase and fixed for 90 min in formalin. Primary Ab was incubated 1/1,000 for 2 h at room temperature (RT), and secondary Ab (Santa Cruz goat anti-rabbit FITC) at 1/500 was incubated at RT for 1 h. Cells were mounted on ∼1.2% agar pads immobilized within a Gene Frame (ABgene) using a 0.13- to 0.17-mm glass coverslip (VWR). Microscopy was performed on an inverted epifluorescence microscope (Zeiss Axiovert 200M) fitted with a Plan-Neofluar objective (Zeiss 100×/1.30 Oil Ph3), a 300-W xenon arc lamp transmitted through a liquid light guide (Sutter Instruments), and a Sony CoolSnap HQ cooled CCD camera (Roper Scientific). All filters were Modified Magnetron ET Sets from Chroma, and details are available on request. Digital images were acquired and analyzed with METAMORPH software (version V.6.2r6).

Proteinase K Treatment of Whole Cells. Cultures of *Bt* (100 mL) were grown in MM with Hep as the sole carbon source to midexponential phase (OD_{600} ∼ 0.6–0.8; monitored using a Biochrom WPA cell density meter). Cells were harvested by centrifugation and washed in 10 mL PBS before being resuspended in 5 mL buffer. The cells were split into four 1-mL aliquots. To three of the aliquots, 2 mg/mL Proteinase K was added and incubated at 37 °C for 1–16 h, and the fourth sample was left as an untreated control also for 16 h. After incubation with the protease, the samples were centrifuged at 5,000 × *g* for 10 min, and the supernatant discarded. The cell pellets were resuspended in 1 mL PBS, and the proteins were precipitated by the addition of 200 μL trichloroacetic acid and incubation on ice for 30 min. The precipitated proteins were pelleted by centrifugation and washed four times in 1 mL ice cold acetone. The protein pellets were resuspended in 250 μL Laemmli buffer and subjected to SDS/PAGE. Gels were transferred to Whatman Protran BA 85 nitrocellulose membrane. Proteins of interest were detected using antisera raised in rats against BT4661, BT4662, or BT4657. The secondary antibody used was a chicken anti-rat conjugated to HRP. Detection was by chemiluminescence using Biorad Clarity Western ECL Substrate.

Coimmunoprecipitation. Immunoprecipitation was carried out using a Pierce coimmunoprecipitation (co-IP) kit according to the manufacturer's instructions. Briefly, cells from 30 mL WT *Bt* grown on MM Hep to midexponential phase were lysed using BugBuster (Novagen), and solubilized extract was precleared with control resin. The total lysed cell material was passed down an anti-BT4661 antibody column (100 μL agarose resin prepared with 500 μg antibody and cross-linked as per the manufacturer's instructions). The column was incubated with gentle rocking at 4 °C for 16 h to pull out BT4661 and any interacting partners from the lysate. The column was then washed extensively with Elution buffer (pH 2.8; containing primary amines) at 4 °C, and material remaining bound to the immobilized antibody was eluted with a low pH buffer. Eluted fractions were neutralized with 1 M Tris, pH 9.5, and anti-BT4661 (raised in rabbit) or anti-BT4659 (raised in rat) antibodies used to probe membranes. Detection was by anti-rabbit or -rat HRP-conjugated secondary.

Whole-Cell Assays. *Bt* was grown in 5 mL MM with 1% (wt/vol) appropriate GAG or glucose as the sole carbon source to midexponential phase in glass test tubes. Cells were harvested by centrifugation at 5,000 × *g* for 10 min at room temperature and washed in 5 mL PBS (pH 7.1) before being resuspended in 1 mL PBS. Washed cells were assayed against 10 mg mL^{-1} of the appropriate GAG at 37 °C for up to 72 h. Assays were analyzed by TLC, and 2 μL each sample was spotted onto silica plates and resolved in butanol:formate:water (4:8:1) buffer. The plates were dried, and the sugars were visualized using diphenylamine stain.

Recombinant Protein Production. Genes were amplified by PCR using the appropriate primers and the amplified DNA cloned in pET28a using NcoI/XhoI or NheI/XhoI restriction sites generating constructs with either N- or C-terminal His6 tags. Recombinant genes were expressed with *Escherichia coli* strains BL21 (DE3) or TUNER (Novagen), containing the appropriate recombinant plasmid, and cultured to midexponential phase in LB supplemented with 50 μg/mL kanamycin at 37 °C and 180 rpm. Cells were then cooled to 16 °C, and recombinant gene expression was induced by the addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside; cells were cultured for another 16 h at 16 °C and 180 rpm. The cells were then centrifuged at 5,000 × *g* and resuspended in 10 mM Hepes, pH 7.5, with 500 mM NaCl before being sonicated on ice. Recombinant protein was then purified by immobilized metal ion affinity chromatography using a cobalt-based matrix (Talon, Clontech) and eluted with 100 mM imidazole. For the proteins to undergo structural studies (BT1596, BT4661, and BT4656), another step of size exclusion chromatography was used using a Superdex 16/60 S200 (GE Healthcare), with 10 mM Hepes, pH 7.5, and 500 mM NaCl as the eluent, and they were judged to be ≥95% pure by SDS/PAGE. Protein concentrations were determined by measuring absorbance at 280 nm using the molar extinction coefficient calculated by ProtParam on the ExPasy server (web.expasy.org/protparam/).

ITC. The affinity of $BT4661^{SGBP}$ and $BT4659^{SusD\text{-}like}$ for oligo- and polysaccharides was quantified by ITC using a Microcal VP-ITC. The protein sample (50 μM), stirred at 300 rpm in a 1.4-mL reaction cell, was injected with 26 × 10-μL aliquots of ligand. Titrations were carried out in 20 mM Na-Hepes buffer, pH 7.5, at 25 °C. Integrated binding heats minus dilution heat controls were fit to a single set of sites binding model to derive K_a, ΔH, and n (number of binding sites on each molecule of protein) using Microcal Origin v7.0. The molar concentration of binding sites present in polysaccharides for each protein was determined by altering the concentration of ligand used for regression of the isotherm until the fit yielded a value of one for n.

SAXS Experiments. Purified $BT4661^{SGBP}$ (as described above) was concentrated to ∼6 mg/mL using centricon devices with a cutoff of

10 kDa. For BT4661 in complex with Hep (BT4661L), 1 mg/mL ligand was added to the protein before concentrating the solution. A dilution series with buffer at 10 mM Hepes, pH 7.5, 250 mM NaCl, and 5% glycerol was prepared with and without 1 mg/mL ligand to obtain five samples covering a concentration range from 0.8 to 5.9 mg/mL for BT4661 in the presence and absence of ligand. The protein samples were filtered through a Millex-GV filter with a 0.22-μm cutoff PVDF membrane before each measurement.

SAXS experiments were carried out at the X33 beamline of the European Molecular Biology Laboratory (Deutsches Electronen Synchrotron) using a Pilatus 500 K detector. A 4.2 mg/mL solution of BSA was measured as a reference and for calibration. The scattering patterns were measured with an exposure time of 4 × 30 s at 288 K. The wavelength was 1.5 Å. The sample to detector distance was set at 2.4 m, leading to scattering vectors q (defined as $q = 4\pi/\lambda\sin\theta$, where 2θ is the scattering angle) ranging from 0.06 to 0.5 $Å^{-1}$. SAXS curves were measured on protein samples with varied concentrations to check for interparticle interactions. Background scattering was measured after each protein sample using the buffer solution and then subtracted from the protein scattering patterns after proper normalization and correction from detector response. Absolute calibration was made with a Lupolen sample. Experiments were carried out at a temperature of 20 °C.

Enzyme Assays. All assays, unless stated, were carried out in 50 mM MES, pH 6.0, or Bis-Tris propane, pH 6.5, with 150 mM NaCl and 5 mM $CaCl_2$ at 37 °C. PL activity was monitored continuously at A_{235} by the formation of the carbon double bond generated through the enzymes beta elimination mechanism of action. $BT4658^{GH88}$ glucuronyl hydrolase activity was determined by monitoring loss of signal at A_{235}. The ability of BT1596 to remove O2 sulfation from unsaturated Hep disaccharides was monitored continuously by the loss of signal at A_{235} by linked assay using $BT4658^{GH88}$ (inactive on O2-sulfated unsaturated Hep disaccharides) to indirectly observe the loss of O2 sulfation. The ability of $BT4656^{6S\text{-}sulf}$ to remove O6 sulfation from GlcNAc6S was monitored by the production of GlcNAc detected by high-pressure anion exchange chromatography (HPAEC) with pulsed amperometric detection using a carbohydrate standard quad waveform for electrochemical detection at a gold working electrode with an Ag/AgCl pH reference electrode and a Carbopac PA-100 guard and analytical column (Dionex; ThermoFisher). The ability of recombinant $BT4654^{ROK}$ to phosphorylate various sugars was evaluated by performing sugar kinase assays as described previously (33). Briefly, ADP generated by the transfer of phosphate to a sugar acceptor is used by pyruvate kinase to generate ATP and pyruvate, which is used by lactate dehydrogenase to oxidize NADH to NAD+, and thus, loss of absorbance at 340 nm can be used to indirectly monitor sugar phosphorylation. Reaction profiles of PL digestion of GAGs were monitored by HPAEC using an AD25 absorbance detector at A_{235} to detect the carbon double-bond products, with H_2O, pH 3.5, as the eluent and a second eluent of H_2O with 3 M NaCl used to generate a linear NaCl gradient to 50% over 80 min. For substrate depletion experiments, the equation $(k_{cat}/K_m)\cdot t = \ln(S_0/S_t)$ was used, and the data were fitted by linear regression. To determine all other kinetic data, the initial rate vs. a range of substrate concentrations were fit to the Michaelis–Menten equation by nonlinear regression using Graphpad Prism 6.0. All assays were run in triplicate unless stated.

Crystallization of BT1596, BT4661, and BT4656. After purification, BT4661 was dialyzed against ultrapure H_2O, whereas BT1596 and BT4656 were carried forward in the same eluent as used for the size exclusion chromatography. All proteins were then concentrated in centrifugal concentrators with a molecular mass cutoff of 30 kDa. Sparse matrix screens were set up in 96-well sitting drop TTP Labtech plates (400-nL drops). Initial BT1596 ligand-bound crystals were obtained at 5 mg/mL with 10 mM ligand in 20% PEG 3350 and KCl. For BT4656, apo- and ligand-bound initial hits were obtained at 20 mg mL^{-1} in 2 M NH_4SO_4, 0.1 M Na citrate, pH 5.6, and 0.2 M Rochelle salts with 20% PEG 3350 and $NaNO_3$, respectively. Apo BT4661 was crystallized at 10 mg mL^{-1} in 18–23% PEG 3350, 350 mM $NaSO_4$, and 0.1 M Bis·Tris Propane, pH 8.0. Truncated BT4661 was crystallized at 10 mg mL^{-1} with 5 mM Hep hexasaccharide in 25% PEG 1500 and 100 mM MMT (1:2:2 – DL-malic acid:MES:Tris base), pH 6. Data were collected at Diamond Light Source (Oxford) on beamlines I02 and I04-1 (0.92 Å) at 100 K. The data were integrated with XDS and scaled with Aimless. Five percent of observations were randomly selected for the Rfree set. The phase problem was solved by molecular replacement using the program Phaser for BT1596 with the Protein Data Bank (PDB) search model ID code 3B5Q and the automated molecular replacement server Balbes for BT4656. The crystal structure of BT4661 was solved using single-wavelength anomalous dispersion (SAD) based on the selenomethionine (SeMet) sites. Structure determination used programs in CCP4i. The sites and phases were determined using the SHELXC/D/E pipeline in CCP4i using intensities (34). Because of the similarity in cell dimensions between the SeMet and the native crystal, the phases from the SAD experiment could be transferred and extended to the higher-resolution native dataset. The phase quality allowed us to use automated model building with ARP/wARP. The models were completed using iterative cycles of refinement with refmac5 and model-building using COOT. The dataset obtained from truncated BT4661 crystallized in presence of ligand was solved with Molrep using the corresponding PDB model from the native apo form. Solvent molecules were added using ARP/wARP solvent and checked manually. The model underwent cycles of model building in COOT and refinement in Refmac5. All other computing used the CCP4 suite of programs. The model was validated using MolProbity, and data statistics, refinement details, and PDB ID codes are reported in Table S2. Structure representations were made using Pymol (version 1.74; Schrödinger).

Site-Directed Mutagenesis. Site-directed mutagenesis was conducted using the PCR-based QuikChange kit (Stratagene) according to the manufacturer's instructions using the appropriate plasmid as the template and appropriate primer pairs. Mutants of $BT4661^{SGBP}$ were run on native PAGE gels in the presence or absence of Hep (0.1% final; added to the gel before polymerization) to assess their ability to bind the GAG.

Comparative Genomics Analysis. PULs similar to *Bt* PUL_{Hep} were searched for in >300 Bacteroidetes genomes. The identification of similar PULs was based on a combination of PUL modular alignments and BlastP searches. Gene composition and order in all of the Bacteroidetes PULs were first computed using the PUL predictor described in PULDB (35). Then, the predicted PULs were aligned to PUL_{Hep} according to their modularity as proposed in the RADS/RAMPAGE method for protein alignment at the domain level (36). Modules taken into account include CAZy families, sensor regulators, sulfatases, and susCD-like genes. Finally, PUL boundaries and limit cases were refined by BLASTP-based analysis at the protein level.

SI Results

SAXS Data Analysis. The values of radii of gyration (R_g) were derived from the Guinier approximation (37): $I(q) = I(0)\exp(-q^2R_g^2/3)$, where $I(q)$ is the scattered intensity, and $I(0)$ is the forward scattered intensity. The radius of gyration and $I(0)$ are inferred from the slope and the intercept, respectively, of the linear fit of $Ln[I(q)]$ vs. q^2 in the q range $q\cdot R_g < 1.3$. The distance distribution function $P(r)$ was calculated on the merged curve by the Fourier inversion of the scattering intensity $I(q)$ using GNOM (38). All scattering curves were indicative of monomeric states of the molecules in solution. The concentration dependence of the R_g value in the low

q region (Table S3), however, was indicative of slight repulsive interactions of the molecules in solution, but otherwise, the curves registered at different concentrations were perfectly superimposable. The lowest concentration curves were considered too noisy and were not further taken into account. The estimated R_g value from the Guinier approximation was, therefore, extrapolated at zero concentration for both samples based on the linear correlation of the four other concentrations. A merged curve using the low concentration curve (1.7 mg/mL) in the low q region (<0.085 $Å^{-1}$) and the high concentration curve (5.9 mg/mL) was produced for each of the proteins BT4661 with and without ligand and used in subsequent steps.

The low-resolution shape of BT4661 and BT4661L in solution was determined ab initio from the merged scattering curve using the program GASBOR (39). This program restores low-resolution shapes of proteins and calculates a volume filled with densely packed spheres (dummy residues of 3.8-Å diameter), fitting the experimental scattering curve by a simulated annealing minimization procedure with a nearest neighbor distribution constraint and leading to χ^2 values of 0.99 and 1.00 for BT4661 and BT4661L, respectively. In comparison, CRYSOL (40) was used to calculate the fit of the crystal structure (PDB ID code 4AK1) to the experimental scattering curve giving χ^2 values of 2.07 and 4.72, respectively.

Several independent fits were run with no symmetry restriction, and the stability and coherence of the solution were checked by calculating normalized spatial discrepancy values of 1.62 (BT4661) and 1.54 (BT4661L) for five independent calculations using DAMAVER (41). The obtained envelopes were compared with the crystal structure by superimposing the objects using PyMOL (version 1.74; Schrödinger).

BT4657^{PL12} + 1 (Active Site) Specificity. When a mixture of two heavily sulfated tetrasaccharides (4 and 5 in Fig. S4*B*) was treated with BT4657^{PL12}, only glycan 4 was digested, producing Δ4,5UA-GlcNS6S and Δ4,5UA2S-GlcNS6S (Fig. S4*C*). This finding suggests that 2-O sulfation is inhibitory to BT4657^{PL12} but leaves ambiguity if it is at the +1 or −2 subsite [the scissile bound is between the sugars at −1 and +1, with the nonreducing and reducing ends of the substrate extending into the negative and positive subsites, respectively (42)]. Treatment of the tetrasaccharide mix with BT1596, a 2-O sulfatase that specifically removes 2-O sulfate groups from the nonreducing end of Δ4,5UA, produced a tetrasaccharide (glycan 6 in Fig. S4*B*) that was still not cleavable by BT4657 (Fig. S4*C*). These data show that BT4657^{PL12} cannot tolerate 2-O sulfation of the UA at its +1 subsite and provide a biochemical basis for its preference for HS and ΔSHep, which contain much lower amounts of 2-O sulfation than Hep (Fig. 1).

Sulfatases Biochemical and Structural Characterization. BT1596$^{2S\text{-}sulf}$ and BT4656$^{6S\text{-}sulf}$ both have a pocket topology with putative metal binding sites located at the base of the pocket. In BT1596$^{2S\text{-}sulf}$, this site comprises D40, H41, D299, and H300, whereas in BT4656$^{6S\text{-}sulf}$, the side chains of D37, D38, D349, and Q350 fulfill this role. In both proteins, the formylglycine residues and the catalytically labile sulfate complete an octa-coordinated geometry of the bound metal. A calcium ion could only be modeled into the site of BT4656$^{6S\text{-}sulf}$. The importance of the metal binding site is shown by the observation that the mutations D37A, D38A, and D349A in BT4656 and H300A in BT1596$^{2S\text{-}sulf}$ completely inactivated the respective enzymes (Table S6). Neither enzyme could be inactivated by EDTA, suggesting that the metal is either bound extremely tightly or inaccessible to EDTA.

In BT1596$^{2S\text{-}sulf}$, the O2 sulfate group of the Δ4,5-unsaturated UA makes potential ionic interactions with the Nζ of K133 and K312 and hydrogen bonds to Nδ of N101 and Nε2 of H196, and it may make polar contact with the backbone amine of the catalytic residue S72. The mutations K133A and N106A caused substantial reductions in catalytic activity, supporting the important role of the lysine and asparagine in sulfate binding (Table S6). E394 interacts with O3 of the unsaturated IdoA, and the E394A variant displayed a 1,000-fold decrease in k_{cat}/K_m. Y395 interacts with the glycosidic oxygen through its side chain. The Y395A mutation did not affect k_{cat}/K_m but greatly increased K_m and k_{cat}. This result suggests that the residue contributes to binding but does not contribute to transition state stabilization, whereas the commensurate increase in k_{cat} likely reflects increased rate of product departure (Table S6). The reducing end GlcNS6S makes few interactions with BT1596$^{2S\text{-}sulf}$, and R245 makes a hydrogen bond to O3 and a possible ionic interaction with the N2 sulfate, both via Nδ2. There is also a potential interaction between the O6 sulfate and Nζ of K133. Mutation of R245 to Ala had the same effect as the Y395A mutation discussed above, showing an important role in substrate binding. The Y395F mutant was completely inactive, suggesting that the more hydrophobic nature of the side chain prevents substrate binding.

In BT4656$^{6S\text{-}sulf}$, the N98A and K362A mutations both inactivate the enzyme (Table S6). N98 sits at the bottom of the active site and interacts with the O6 sulfate via Nδ2. K362 potentially hydrogen bonds to the thioester linkage of the O6 sulfate but also interacts with D38 of the calcium binding site. Mutation of K362 could potentially disrupt the metal binding site and thus, the loss of activity of the K362A mutant. K125 seems to play an important role in O6 sulfate binding, because K125A substitution resulted in a 1,000-fold reduction in activity. Q147 binds to the endocyclic ring oxygen and the O6 sulfate via Nε2, and the Q147A variant causes around a 30-fold reduction in activity (Table S6). The importance of the interactions between E446 and O1 and N2 of the substrate is illustrated by the 1,000-fold decrease in activity mediated by the E446A mutation. The H447A substitution decreased activity 500-fold, suggesting that H447 is the only residue making productive interactions via O3.

Both BT1596 and BT4656 have predicted signal peptides (www.cbs.dtu.dk/services/LipoP/), suggesting a periplasmic location for these enzymes, consistent with the biochemical data. The likely sulfamidase BT4655 is a predicted TAT protein (www.cbs.dtu.dk/services/TatP/) and therefore, also likely periplasmic. Note that the N-terminal methionine in each of these proteins is incorrect in the database sequence and should be 23 residues downstream in BT4656 (correct *N* term starts MKSN), 7 residues upstream in BT1596 (correct *N* term starts MKTI), and 37 residues upstream in BT4655 (correct *N* term starts MDRR).

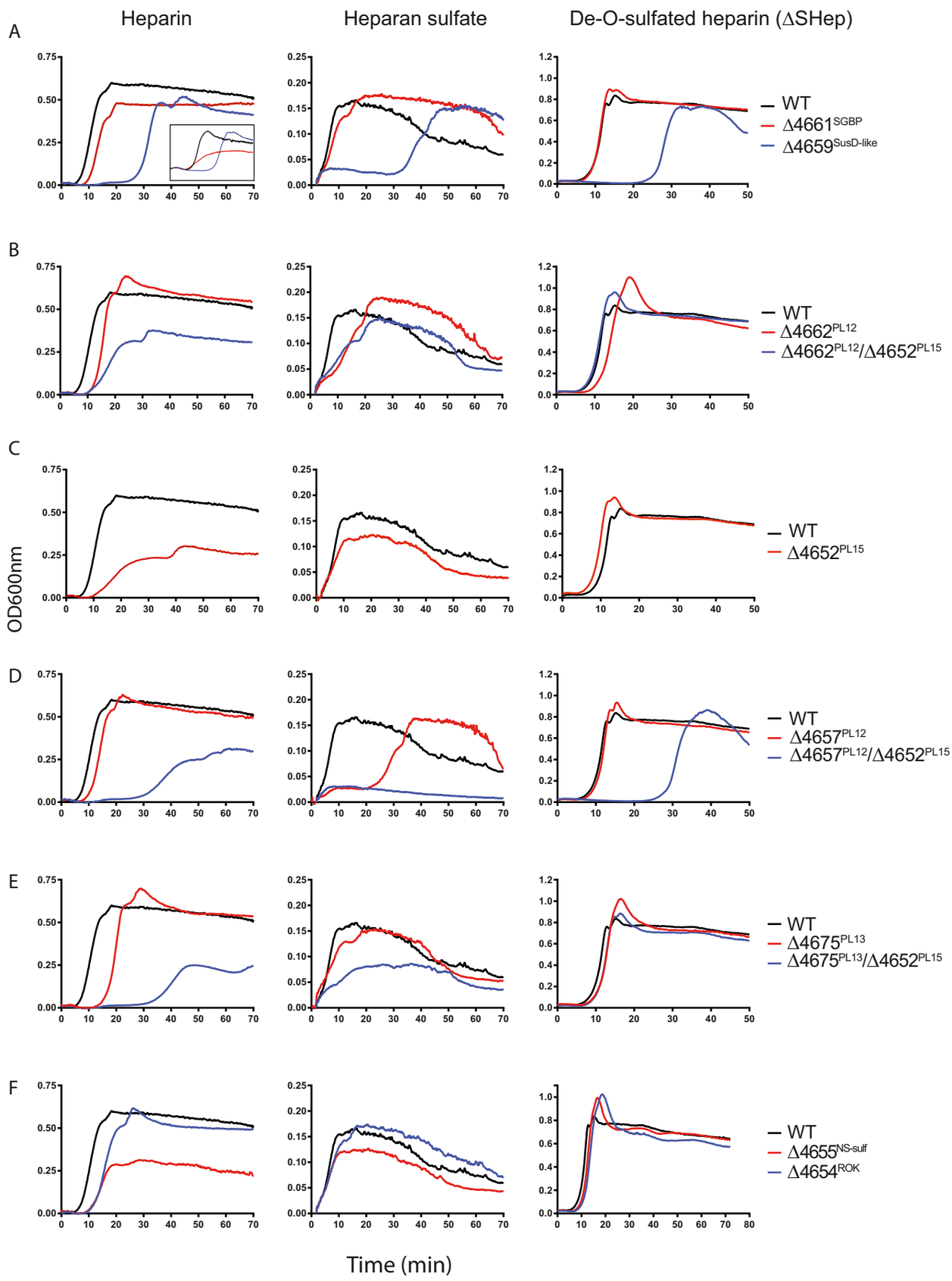

Fig. S1. Growth of WT and mutant strains of *Bt* on Hep, HS, and ΔSHep. (*A*) Mutant strains with deletions in PULHep glycan-binding proteins BT4661 and BT4659. *Left*, *Inset* shows growth on heparin oligosaccharides. (*B*) BT4662 PL12 single- and BT4662 PL12/BT4652 PL15 double-mutant strains. (*C*) BT4652 PL15 mutant strain. (*D*) BT4657 PL12 single- and BT4657/BT4652 PL15 double-mutant strains. (*E*) BT4675 PL13 single- and BT4675 PL13/BT4652 PL 15 double-mutant strains. (*F*) BT4655 NS sulfatase and BT4654 GlcNAc kinase mutant strains. All growths were performed in MM supplemented with 10 mg/mL appropriate polysaccharide. OD_{600} measurements were taken every 20 min after bacterial inoculation using a plate reader.

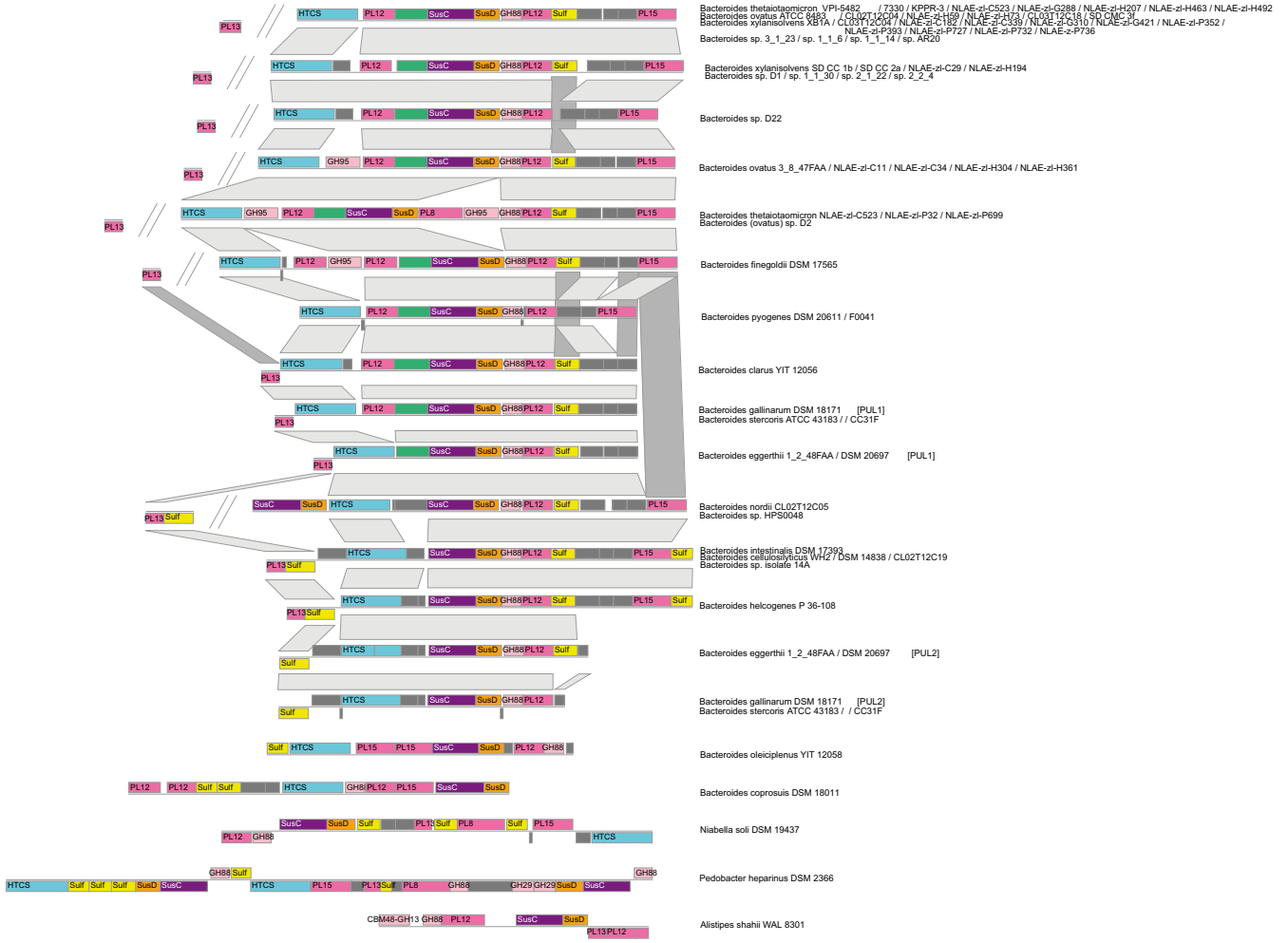

Fig. S2. PUL_{Hep} distribution and synteny. Genes are represented above or under the scaffold (bold line) to distinguish the coding strain. High distance between the left-most PL13-coding gene, separated by 5–20 genes from the main PUL region, is depicted by a double slash interrupting the scaffolds. Homologous proteins across species with similar PUL modular arrangement are connected by gray trapezoids. Species names are indicated on the right. PUL 1 and 2 are used to indicate discrete but similar PULs within the same species.

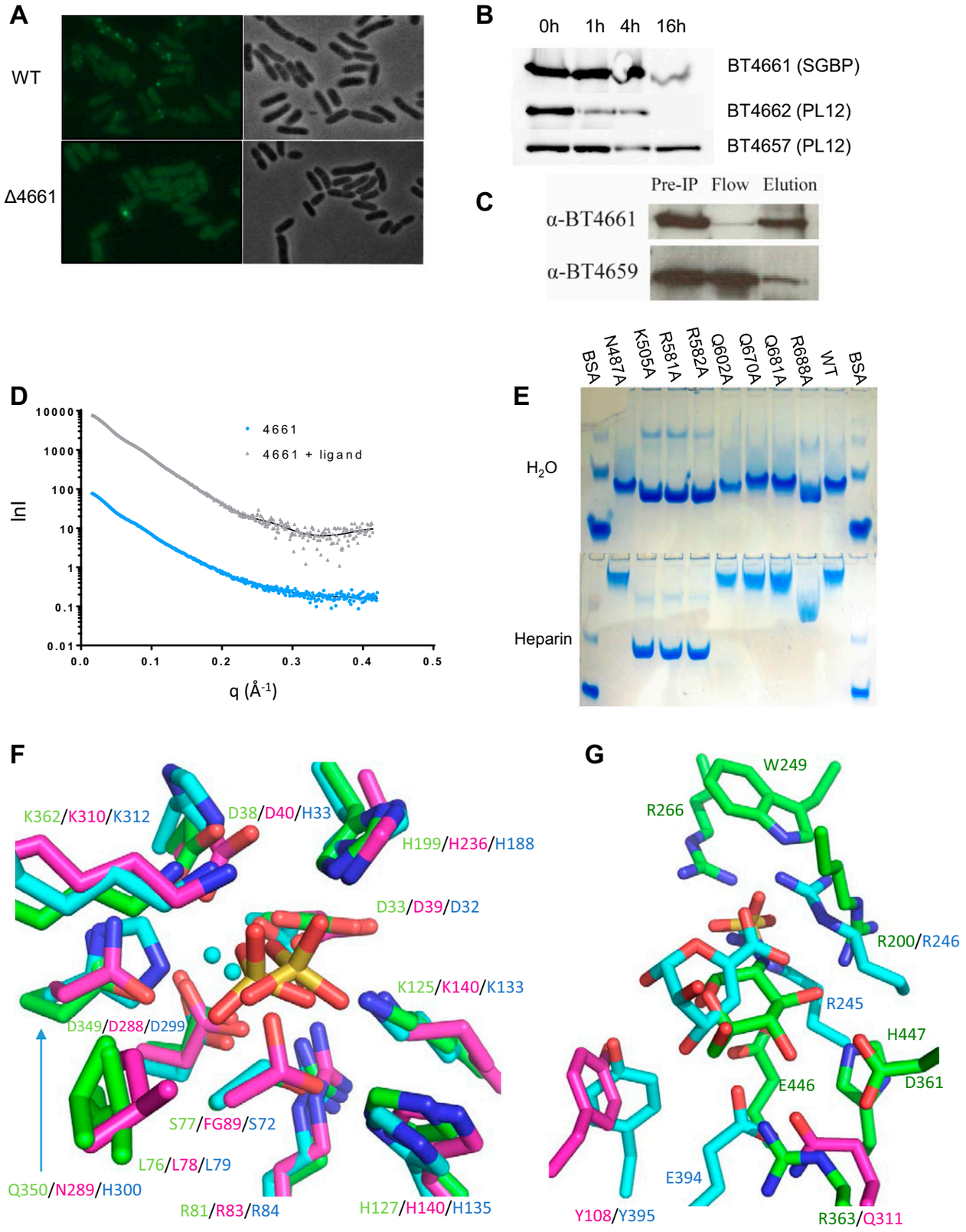

Fig. S3. Cellular localization of proteins encoded by PUL_{Hep}, SAXS curves, and key binding residues of $BT4661^{SGBP}$ and comparison of the active sites of the sulfatases with structural homologs. (*A*) Immunofluorescence labeling of Glc-grown WT or ΔBT4661 cells with (*Left*) anti-4661 antibody or (*Right*) phase contrast. (*B*) Western blots showing Hep-grown whole cells treated with proteinase K and probed with polyclonal antibodies raised against recombinant forms of $BT4661^{SGBP}$, $BT4662^{PL12}$, and $BT4657^{PL12}$. Proteinase K cannot cross the outer membrane and therefore, can only act on surface-located proteins, revealing that $BT4661^{SGBP}$ and $BT4662^{PL12}$ are surface-located, whereas $BT4657^{PL12}$ is most likely periplasmic. (*C*) Co-IP of $BT4661^{SGBP}$ and $BT4659^{SusD\text{-}like}$ from lysate of Hep-grown cells using anti-BT4461 polyclonal antibody immobilized to agarose beads. IP, immunoprecipitation. (*D*) Experimental SAXS curves and the fitted scattering profiles calculated by GASBOR. Experimental merged data are shown as gray triangles (liganded $BT4661^{SGBP}$) or blue circles (apo $BT4661^{SGBP}$). The scattering curves fitted with GASBOR are shown as solid lines. lnI, logarithmic representation of intensity; *q*, scattering angle in angstroms^{-1}. (*E*) Affinity gel electrophoresis of $BT4661^{SGBP}$ alanine mutants of residues that interact with the fully sulfated Hep-derived hexasaccharide ligand. *Upper* shows native gel with no added ligand, and *Lower* shows native gel with added Hep (0.1% final). Loss of retardation of K505A, R581A, and R582A on the Hep gel reveals that these residues are critical in ligand recognition. (*F*) Overlay of $BT4656^{6S\text{-}sulf}$ (green), human galactosamine-6-sulfatase (PDB ID code 4FDI; magenta), and $BT1596^{2S\text{-}sulf}$ (cyan), showing the conservation in the catalytic center. The sulfate groups shown are bound to $BT4656^{6S\text{-}sulf}$ and $BT1596^{2S\text{-}sulf}$, whereas the calcium ions are bound to $BT4656^{6S\text{-}sulf}$ and 4FDI. Note that residue 89 in 4FDI is formyl glycine (FG). (*G*) Overlay of $BT4656^{6S\text{-}sulf}$ with GlcNS6S bound (green), 4FDI (magenta), and $BT1596^{2S\text{-}sulf}$ with 4,5UA2S-GlcNS6S bound (cyan), showing variability in the glycone binding region. For clarity, only the sugar interacting with the protein has been shown.

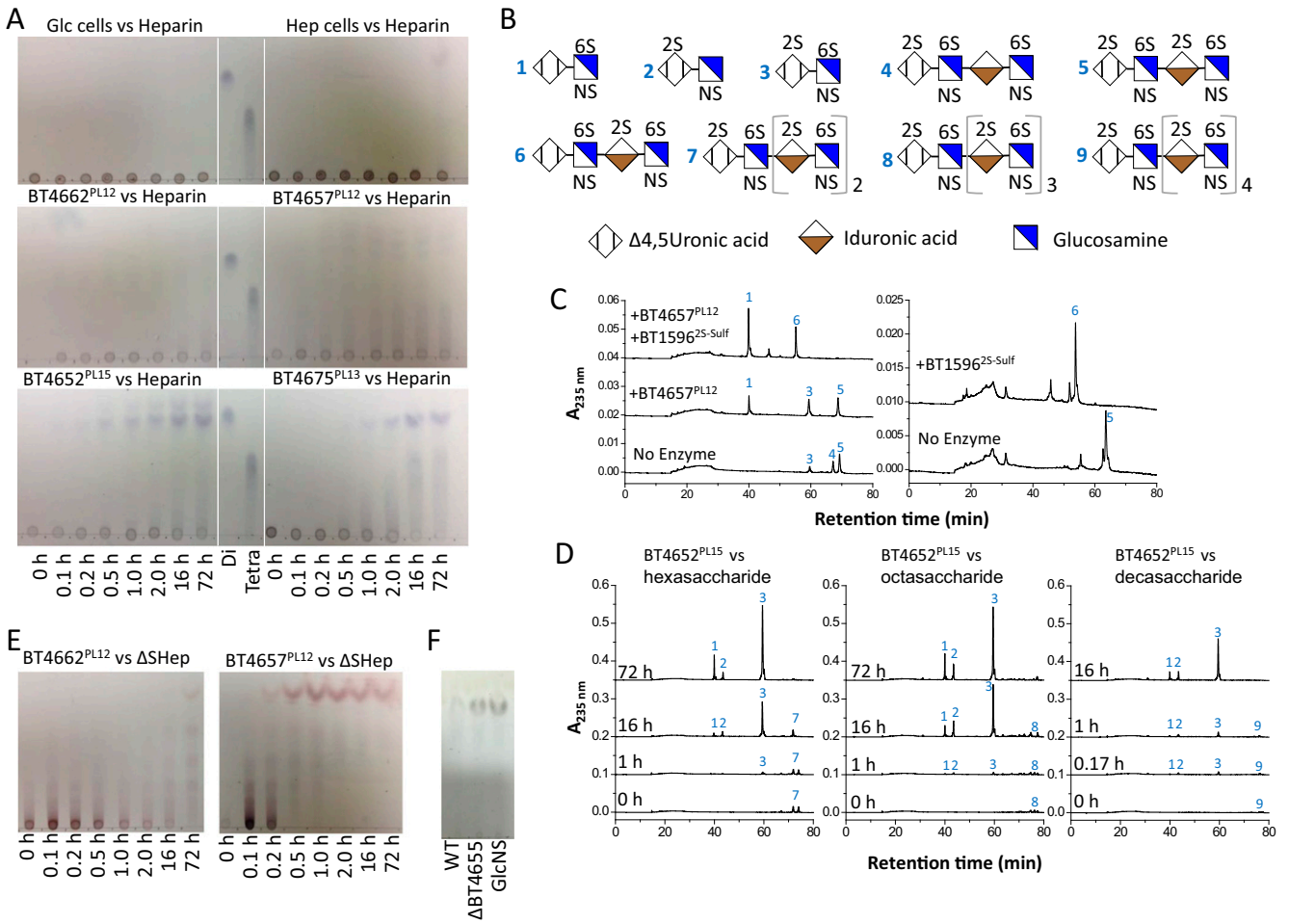

Fig. S4. Activity of whole cells and recombinant PUL_{Hep} enzymes against Hep and Hep oligosaccharides and evidence that BT4655 is a sulfaminidase. (*A*) Activity of whole cells and recombinant lysases (0.1 μM) against Hep (10 mg/mL). *Top* shows cells grown on Glc or Hep as the sole carbon source to midexponential phase, washed, and assayed against Hep to determine surface enzyme activity. *Middle* and *Bottom* are products released by recombinant forms of the different PUL-encoded lyases against Hep. Di and Tet are Hep disaccharide and tetrasaccharide standards, respectively (structures 2 and 5, respectively, in *B*). (*B*) Identity of the sugars labeled in *C* and *D*. (*C*) Chromatograms showing sulfate tolerances of $BT4657^{PL12}$ against 4, a sulfated tetrasaccharide lacking a single O2 sulfation, and 5, a fully sulfated tetrasaccharide, pre- and postaddition of $BT1596^{2S\text{-}sulf}$. (*D*) Chromatograms showing exoprocessivity of $BT4652^{PL15}$ against Hep oligosaccharides (all at 50 μM substrate). (*E*) TLC analysis of $BT4662^{PL12}$ and $BT4657^{PL12}$ product profiles against ΔSHep (10 mg/mL). (*F*) Supernatants of stationary-phase WT and $\Delta BT4655^{NS\text{-}Sulf}$ cells grown on Hep. GlcNS lane is a standard.

Table S1. ITC data for BT4659$^{SusD\text{-}like}$ and BT4661^{SGBP}

Ligand	Protein	$K_a \times 10^4$, M^{-1}	ΔH, kcal·mol^{-1}	TΔS, kcal·mol^{-1}	*n*
Polysaccharides					
HS	BT4659$^{SusD\text{-}like}$	1.4 (±0.6)	−14.0 (±0.3)	−8.3	1.0 (±0.01)
	BT4661^{SGBP}	27.0 (±4.0)	−9.2 (±0.1)	−1.8	1.0 (±0.04)
Hep	BT4659$^{SusD\text{-}like}$	1.5 (±0.2)	−21 (±5.0)	−15.3	0.8 (±0.1)
	BT4661^{SGBP}	40.0 (±9.0)	−9.7 (±0.3)	−2.1	1.0 (±0.02)
ΔSHep	BT4659$^{SusD\text{-}like}$	2.1 (±0.4)	−9.6 (±0.2)	−3.7	0.9 (±0.2)
	BT4661^{SGBP}	11.0 (±8.1)	−4.6 (±1.5)	2.3	1.0 (±0.01)
Hep oligosaccharides					
Dp2	BT4659$^{SusD\text{-}like}$	NB			
	BT4661^{SGBP}	NB			
Dp4	BT4659$^{SusD\text{-}like}$	NB			
	BT4661^{SGBP}*	8.5 (±0.4)	−6.5 (±0.1)	0.2	0.8 (±0.01)
Dp6	BT4659 $^{SusD\text{-}like}$	NB			
	BT4661^{SGBP}*	31.0 (±2.0)	−8.3 (±0.1)	−0.8	0.8 (±0.01)
Dp8	BT4659$^{SusD\text{-}like}$	NB			
	BT4661^{SGBP}	37.0 (±1.0)	−8.7 (±0.2)	−1.1	0.7 (±0.01)
Dp10	BT4659$^{SusD\text{-}like}$	NB			
	BT4661^{SGBP}*	55.0 (±3.0)	−13.0 (±0.1)	−5.2	0.5 (±0.01)

Data are averages and SDs of at least two replicates, except where indicates. NB, no binding at the maximum concentration tested (1 mM for oligosaccharides).

*Data are from a single titration with errors of fit.

Table S2. Crystal data

Parameter	BT4661 SeMet	BT4661	BT4661D5-D6-hexasaccharide	BT1596-Apo	BT1596-UA2SGlcNS6S	BT4656-GlcNS6S
Data collection	01/08/09	15/05/09	27/09/09	22/04/14	10/07/14	04/08/14
Beamline	Diamond light source I04	Diamond light source I02	Diamond light source I02	Diamond light source I04-1	Diamond light source I04-1	Diamond light source I04-1
Cell parameters	a = 158.78, b = 158.678, c = 137.45; $\alpha = \beta = 90°$, $\gamma = 120°$	a = 158.62, b = 158.63, c = 137.06; $\alpha = \beta = 90°$, $\gamma = 120°$	a = 52.6, b = 68.15, c = 85.14; $\alpha = \beta = \gamma = 90°$	a = 68.87, b = 68.87, c = 110; $\alpha = \beta = \gamma = 90°$	a = 69.21, b = 114.535, c = 127.29; $\alpha = \gamma = 90°$, $\beta = 100°$	a = 71.22, b = 74.55, c = 95; $\alpha = \beta = \gamma = 90°$
Space group	$P6_322$	$P6_322$	$P2_12_12_1$	$P4_1$	$P2_1$	$P2_12_12_1$
Wavelength (Å)	0.98	0.98	0.98	0.92	0.92	0.92
Resolution (Å)	51.30–3.10 (3.27–3.10)	45.79–1.95 (2.06–1.95)	29.77–1.35 (1.42–1.35)	44.53–1.43 (1.45–1.43)	46.40–1.90 (1.93–1.90)	47.50–1.39 (1.41–1.39)
Rmerge	19.4 (44.7)	0.10 (0.36)	0.04 (0.39)	0.06 (0.72)	0.10 (0.74)	0.07 (0.45)
Mean I/σI	19.6 (11.0)	10.5 (3.9)	21.8 (4.2)	15.1 (1.6)	9.0 (2.6)	22.8 (5.8)
CC 1/2	N/A	N/A	N/A	0.999 (0.908)	0.996 (0.761)	0.995 (0.745)
Completeness	100 (100)	99.5 (99.5)	98 (97.3)	100 (99.9)	(99.6) (76.1)	99.5 (94.7)
Anomalous completeness	100 (100)	—	—	—	—	—
Redundancy	43.4 (45.1)	4.3 (4.3)	6.6 (6.4)	7.3 (4.7)	3.8 (3.9)	7.4 (6.7)
Anomalous redundancy	23.5 (23.8)	—	—	—	—	—
Total reflection	797,932	316,307	440,451	689,619 (21,786)	587,944 (29,303)	751,944 (31,909)
Total unique reflection	18,386	73,908	66,383	94,403 (4,683)	153,578 (7,607)	101,665 (4,749)
Structure refinement						
Rwork/Rfree	N/A	0.22/0.25	0.16/0.19	0.13/0.16	0.16/0.20	0.11/0.13
rmsd	N/A					
Bond lengths	N/A	0.010	0.013	0.011	0.012	0.011
Bond angles	N/A	1.21	1.60	1.56	1.53	1.53
Ramachandran plot						
Favored (%)	N/A	98.3	97.6	97.4	97.4	96.2
Allowed (%)	N/A	99.8	100	99.8	99.8	99.8
Outliers	N/A	0.2	0.0	0.2	0.2	0.2
Mean B factor						
Wilson	N/A	17.5	16.42	16.51	17.62	9.10
Main chain	N/A	29.33	16.48	18.78	18.88	9.87
Side chain	N/A	28.44	21.43	21.19	21.36	11.57
Ligand/water	N/A	–/27.05	23.6/28.9	26.48/37.39	40.89/30.62	6.1/25.21
PDB ID code	N/A	4AK1	4AK2	5G2U	5G2T	5G2V

N/A, not applicable.

Table S3. Experimental radius of gyration (R_g), forward scattered intensity [$I(0)$], and maximal distances (D_{max}) values extracted from the SAXS curves of BT4661 and BT4661 + ligand (BT4661L)

Protein	R_g or extrapolated R_g, Å	M_w from I_0	$I(0)$ or R_g from $P(r)$	D_{max} from $P(r)$, Å
[BT4661] (mg/mL)				
0.75	4.33		92	
1.48	4.33		90	
2.87	4.15		90	
4.54	4.05		86	
5.77	4.00		79	
Bt4661	4.43	78.8	4.58	150
[BT4661L] (mg/mL)				
0.78	4.40		96	
1.67	4.30		90	
3.09	4.12		90	
4.52	4.08		93	
5.97	3.98		90	
Bt4661L	4.43	78.8	4.43	150

Table S4. Catalytic activities of the PUL_{Hep}-encoded PLs

Substrate and enzyme	K_m, mg·mL or μM	k_{cat}, min^{-1}	k_{cat}/K_m, min^{-1} mg^{-1}·mL or min^{-1} M^{-1}
Hep*			
$BT4662^{PL12}$	>8	—	325 (±28)
$BT4657^{PL12}$	<0.5	2,389 (±463)	—
$BT4675^{PL13}$	<0.1	4,447 (±54)	—
$BT4652^{PL15}$	0.68 (±0.20)	500 (±49)	787 (±234)
HS*			
$BT4662^{PL12}$	>2	—	819 (±85)
$BT4657^{PL12}$	0.036 (±0.011)	1,193 (±167)	33,139 (±9,842)
$BT4675^{PL13}$	<0.1	3,587 (±68)	—
$BT4652^{PL15}$	>2.0	—	740 (±48)
ΔSHep*			
$BT4662^{PL12}$	>8	—	120 (±13)
$BT4657^{PL12}$	0.079 (±0.002)	1,856 (±167)	23,512 (±2,514)
$BT4675^{PL13}$	NA	NA	NA
$BT4652^{PL15}$	>8	—	0.92 (±0.09)
Hep DP = 4			
$BT4662^{PL12}$	NA	NA	NA
$BT4657^{PL12}$	NT	NT	NT
$BT4675^{PL13}$	—	108 (±15)	—
$BT4652^{PL15\dagger}$	4.4 (±0.5)	1,023 (±158)	4.9×10^8 ($\pm 1.3 \times 10^7$)
Hep DP = 6			
$BT4662^{PL12}$	NA	NA	NA
$BT4657^{PL12}$	NT	NT	NT
$BT4675^{PL13}$	—	1,904 (±361)	—
$BT4652^{PL15}$	16.5 (±3.1)	2,316 (±186)	1.5×10^8 ($\pm 2.2 \times 10^7$)
Hep DP = 8			
$BT4662^{PL12}$	NA	NA	NA
$BT4657^{PL12}$	NT	NT	NT
$BT4675^{PL13}$	—	890 (±214)	—
$BT4652^{PL15\dagger}$	18.6 (±1.2)	3,349 (±56)	1.9×10^8 ($\pm 1.1 \times 10^7$)
Hep DP = 10			
$BT4662^{PL12}$	—	—	1.5×10^6 ($\pm 3.1 \times 10^5$)
$BT4657^{PL12}$	NT	NT	NT
$BT4675^{PL13}$	NT	NT	NT
$BT4652^{PL15\dagger}$	≥70	—	3.5×10^7 ($\pm 1.2 \times 10^6$)

Buffer composition was 50 mM MES, pH 6.0, with 150 mM NaCl and 2 mM $CaCl_2$. NA, no activity detected; NT, substrate was not tested.

*Polysaccharide data are in milligrams per milliliter.

†Experiments were technical duplicates.

Table S5. Lyase end point assays

Enzyme(s)	Hep, mM/mg (%)	HS, mM/mg (%)	ΔSHep, mM/mg (%)
$BT4662^{PL12}$	0.1 ± 0.004 (5)	1.0 ± 0.022 (38)	3.1 ± 0.08 (82)
$BT4657^{PL12}$	0.5 ± 0.01 (27)	1.4 ± 0.02 (54)	3.7 ± 0.06 (100)*
$BT4675^{PL13}$	1.7 ± 0.11 (94)	1.5 ± 0.04 (56)	N/A
$BT4652^{PL15}$	1.7 ± 0.06 (95)	1.5 ± 0.002 (58)	Incomplete
$BT4652^{PL15}$/$BT4675^{PL13}$	1.8 ± 0.01 (99)	1.9 ± 0.02 (71)	—
$BT4652^{PL15}$/$BT4657^{PL12}$	1.7 ± 0.01 (99)	2.4 ± 0.02 (91)	—
$BT4675^{PL13}$/$BT4657^{PL12}$	1.6 ± 0.07 (93)	2.1 ± 0.11 (80)	—
$BT4652^{PL15}$/$BT4662^{PL12}$	1.7 ± 0.02 (98)	2.5 ± 0.22 (93)	—
$BT4662^{PL12}$/$BT4675^{PL13}$	1.4 ± 0.04 (77)	2.2 ± 0.04 (84)	—
$BT4662^{PL12}$/$BT4657^{PL12}$	0.5 ± 0.01 (29)	1.5 ± 0.01 (59)	—
$BT4652^{PL15}$/$BT4675^{PL13}$/$BT4657^{PL12}$	1.7 ± 0.01 (95)	2.4 ± 0.02 (92)	—
$BT4652^{PL15}$/$BT4675^{PL12}$/$BT4662^{PL12}$	1.7 ± 0.05 (95)	2.2 ± 0.04 (86)	—
$BT4652^{PL15}$/$BT4657^{PL12}$/$BT4662^{PL12}$	1.8 ± 0.01 (100)*	2.6 ± 0.03 (100)*	—
$BT4675^{PL13}$/$BT4657^{PL12}$/$BT4662^{PL12}$	1.4 ± 0.02 (82)	2.3 ± 0.02 (89)	—
$BT4652^{PL15}$/$BT4657^{PL12}$/$BT4675^{PL13}$/$BT4662^{PL12}$	1.7 ± 0.01 (95)	2.5 ± 0.05 (97)	—

N/A, not active.

*100% is the maximum digestion observed with any lyase combination for that particular GAG and was then used to normalize other values against the same GAG. Buffer composition was 50 mM MES, pH 6.0, with 150 mM NaCl and 2 mM $CaCl_2$.

Table S6. Catalytic activities of the PUL_{Hep}-encoded *O*-sulfatases, the GH88, and ROK kinase

Substrate	Enzyme	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m ($min^{-1}\cdot M^{-1}$)
GH88				
UA-GlcNS*	BT4658	0.1 (±0.002)	2,211 (±600)	1.8×10^7 ($\pm 4.6 \times 10^6$)
UA-GlcNAc6S*	BT4658	>0.1	—	1.7×10^7 ($\pm 5.1 \times 10^6$)
UA-GlcNAc*	BT4658	>0.15	—	5.0×10^6 ($\pm 2.3 \times 10^5$)
2O-Sulfatase				
UA2S-GlcNAc6S	BT1596	>100	—	5.5×10^5 ($\pm 1.3 \times 10^5$)
UA2S-GlcNAc	BT1596	>150	—	3.6×10^5 ($\pm 9.5 \times 10^4$)
UA2S-GlcNS6S	BT1596†	0.4 (±0.05)	198 (±34)	5.2×10^6 ($\pm 1.2 \times 10^6$)
UA2S-GlcNS6S	$BT1596^{H33A}$			1.0×10^3 ($\pm 5.4 \times 10^1$)
UA2S-GlcNS6S	$BT1596^{N101A}$			6.6×10^3 ($\pm 4.9 \times 10^2$)
UA2S-GlcNS6S	$BT1596^{K133A}$			1.6×10^4 ($\pm 2.6 \times 10^3$)
UA2S-GlcNS6S	$BT1596^{R245A}$			2.2×10^6 (± 1.4 x10^5)
UA2S-GlcNS6S	$BT1596^{R246A}$			2.1×10^3 ($\pm 6.8 \times 10^1$)
UA2S-GlcNS6S	$BT1596^{E394A}$			4.3×10^3 ($\pm 5.7 \times 10^2$)
UA2S-GlcNS6S	$BT1596^{Y395A}$			1.2×10^6 ($\pm 1.0 \times 10^5$)
6O-Sulfatase				
GlcNAc6S	BT4656†	—	—	2.0×10^5 ($\pm 4.0 \times 10^4$)
GlcNAc6S	$BT4656^{K125A}$			1.6×10^2 ($\pm 1.3 \times 10^1$)
GlcNAc6S	$BT4656^{Q147A}$			6.9×10^3 ($\pm 4.6 \times 10^2$)
GlcNAc6S	$BT4656^{R200A}$			9.8×10^4 ($\pm 6.1 \times 10^3$)
GlcNAc6S	$BT4656^{R363A}$			2.0×10^2 ($\pm 9.9 \times 10^1$)
GlcNAc6S	$BT4656^{E446A}$			1.1×10^3 ($\pm 8.0 \times 10^1$)
GlcNAc6S	$BT4656^{H447A}$			2.6×10^2 ($\pm 2.3 \times 10^1$)
ROK				
GlcN	BT4654	0.2 (±0.02)	4,581 (±684)	3.1×10^7 ($\pm 3.9 \times 10^6$)
GlcNAc	BT4654	4.1 (± 0.4)	3,449 (±895)	8.7×10^5 ($\pm 3.1 \times 10^5$)
GlcNS	BT4654	—	—	3.4×10^1 (±1.1)
Glucose	BT4654	114 (±5)	1,296 (±238)	1.1×10^4 ($\pm 2.1 \times 10^3$)
Mannose	BT4654	37 (±9)	867 (±115)	2.4×10^4 ($\pm 2.9 \times 10^3$)
ManN	BT4654	19 (±4)	33 (±2)	6.2×10^2 ($\pm 4.1 \times 10^1$)
ATP‡	BT4654	0.9 (±0.3)	1,954 (±114)	2.4×10^6 ($\pm 5.7 \times 10^5$)

*The GH88 displayed no activity against the UA2S versions of these disaccharides.

†WT BT1596 and BT4656 sulfatases are actually S72C and S77C mutants, respectively, because *E. coli* can only convert cysteine to the formylglycine required for activity. Note that the S72A, H196A, D299A, H300A, and Y395F mutants of BT1596 and the S77A, D37A, D38A, N98A, H199A, D349A, D361A, and K362A mutants of BT4656 displayed no detectable activity. In addition, the H135A and K312A mutants of BT1596 and the H127A, W249A, R266A, and Q350A mutants of BT4656 were made, but recombinant forms failed to express.

‡Values were determined using GlcNAc as the saturating sugar at concentration of 50 mM. The buffer used was 100 mM Tris·HCl, pH 8.0.